Quantitative lung CT analysis for the study and diagnosis of Chronic Obstructive Pulmonary Disease

Felix J.S. Bragman

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Centre for Medical Image Computing

Department of Medical Physics & Biomedical Engineering

University College London

March 16, 2018

I, Felix J.S. Bragman, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

The importance of medical imaging in the research of Chronic Obstructive Pulmonary Disease (COPD) has risen over the last decades. COPD affects the pulmonary system through two competing mechanisms; emphysema and small airways disease. The relative contribution of each component varies widely across patients whilst they can also evolve regionally in the lung. Patients can also be susceptible to exacerbations, which can dramatically accelerate lung function decline. Diagnosis of COPD is based on lung function tests, which measure airflow limitation. There is a growing consensus that this is inadequate in view of the complexities of COPD. Computed Tomography (CT) facilitates direct quantification of the pathological changes that lead to airflow limitation and can add to our understanding of the disease progression of COPD.

There is a need to better capture lung pathophysiology whilst understanding regional aspects of disease progression. This has motivated the work presented in this thesis. Two novel methods are proposed to quantify the severity of COPD from CT by analysing the global distribution of features sampled locally in the lung. They can be exploited in the classification of lung CT images or to uncover potential trajectories of disease progression. A novel lobe segmentation algorithm is presented that is based on a probabilistic segmentation of the fissures whilst also constructing a groupwise fissure prior. In combination with the local sampling methods, a pipeline of analysis was developed that permits a regional analysis of lung disease. This was applied to study exacerbation susceptible COPD. Lastly, the applicability of performing disease progression modelling to study COPD has been shown. Two main subgroups of COPD were found, which are consistent with current clinical knowledge of COPD subtypes.

This research may facilitate precise phenotypic characterisation of COPD from CT, which will increase our understanding of its natural history and associated heterogeneities. This will be instrumental in the precision medicine of COPD.

Acknowledgements

I would like to begin by giving my special thanks to Prof. David Hawkes. He provided valuable insight and encouragement over the course of my studies. I am incredibly grateful for his support, especially at the end of my PhD. I am also grateful that he gave me so much academic freedom during my work.

I am especially grateful to Dr. John Hurst for his supervision and support. I am extremely appreciative of his involvement during the project and for his willingness to try and understand my abstract ideas. Importantly, I am indebted to the patience he showed me throughout my PhD.

My gratitude also goes to Dr. Jamie McClelland. He provided me with valuable guidance and I am particularly thankful for his support when results were not positive. I am very appreciative that he always made time to meet with me and importantly, for the moral support he provided throughout my PhD.

I would like to offer my thanks to Dr. Joseph Jacob, who sacrificed many weekends to support my work. I am especially grateful for the help provided by Dr. Tom Doel who has been so generous with his time. I have received advice, help and support from various other colleagues including, but not limited to, Dr. Jorge Cardoso, Dr. Bjoern Eiben and Dr. Marc Modat.

I would like to thank all my friends who have supported me. To Alex, Eliza, Mark and Zach - I would like to thank you for putting up with my moaning, for your support and for sharing so much coffee with me. To Morten, Patrick and Roman, for being such valuable distractions and such great friends.

It is impossible to forget the contribution of my cousin, Connor, my sister and brother, Claudia and Saul, my parents, Keith and Jessica and my girlfriend, Christine. They have seen me at my worst and stood by me through difficult times during my degree and my PhD. I am grateful for their unwavering support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem statement

Chronic Obstructive Pulmonary Disease (COPD) is defined as the progressive development of airflow limitation that is not fully reversible. Airflow is inhibited by two major components; emphysema and small airways disease. The relative contribution of each component varies widely across subjects, demonstrating an important subject-specific variability in disease progression [165]. COPD affects the respiratory system in a heterogeneous fashion, leading to variety of clinical symptoms. An example of this heterogeneity is exacerbation susceptibility in COPD. Exacerbations are acute events where lung function can significantly worsen and may lead to hospitalisation of the patient. Importantly, there is a subgroup of patients who are susceptible to these episodes.

The severity of COPD is defined by the degree of lung function impairment, which is best known as the Forced Expiratory Volume in 1 second (FEV_1). This is currently the most important diagnostic marker for determining the presence of COPD and evaluating response to treatment both in routine clinical care and drug development. However, it is now clearer than ever that the reliance on FEV_1 is inadequate in view of the complexities of COPD. There are multiple phenotypes of COPD with differing clinical characteristics and varying contributions of emphysema and airway disease [85, 231]. There are various subtypes of emphysema that contribute differently to airflow limitation [34]. Pathological alterations in tissue structure can vary regionally with differing affects on lung physiology [104]. These complexities also translate to the molecular level with numerous inflammatory markers that may be responsible for the systemic inflammation of COPD [85]. These difficulties are motivating the development of novel computational tools that enable more precise stratification of patients to facilitate prognosis and cater therapy to the individual.

The importance of medical imaging for researching COPD has steadily risen over the last decades. Computed Tomography (CT) is presently the gold standard for analysing COPD. Through various technological developments, three-dimensional scans that have a high resolution and excellent contrast can now be acquired. This has promoted the development of a variety of algorithms and computational tools that can support the clinical research of COPD to improve diagnosis and management.

A significant amount of high-dimensional data can now be extracted, which is motivating the development of various analytical tools that are both automated and objective in their quantification of disease. A variety of anatomical structures are affected in COPD such as the parenchyma, the airway tree, the lobar fissures and the vasculature. It is important to accurately segment these structures whilst also effectively capturing the disease processes that affect them. The lung anatomy is incredibly complex with large variability across the population. Combined with the pathological alterations that occur, algorithms that aim to segment the pulmonary structures must be flexible. They need to be able to adapt to cases where pathology alters and occludes the structure of interest but also where the lung anatomy differs significantly from the norm.

Methods that can effectively capture regional aspects of disease and other complex representations are necessary. The lung parenchyma can be pathologically altered in a variety of ways. The traditional method for the analysis of lung tissue and the most extensively used in large clinical studies is the global average. Whilst it is clinical intuitive, it may not be specific in view of the complexities of COPD. Novel algorithms and imaging features that are designed specifically to track various aspect of COPD pathology at a local and global scale are needed. Importantly, there must be a delicate balance between interpretability of the results and complexity of the modelling. It may be desirable for a classifier to learn an abstract feature set that leads to accurate results in a classification setting. However, the strength and clinical utility of a diagnostic algorithm is also dependent on the ease of interpretation of the results, that will invariably further our knowledge of COPD.

1.2 Main contributions

The purpose of the research described within this thesis was to develop novel tools for the advanced analysis of lung CT scans to probe various clinical hypotheses related to COPD progression. The main contributions of this thesis are:

- A novel framework for analysing the distribution of density-based and biomechanical features. Local feature distributions are analysed throughout the lung. Their global distribution can be exploited to study differences between patients and to train a classifier. This algorithm is applied to classify exacerbation susceptible and non-susceptible patients at equal levels of COPD severity. It is tested on three cohorts and is the first example in the literature attempting this problem (**Chapter 3**).

- A novel unsupervised lobe segmentation algorithm. This includes a probabilistic segmentation of the fissures based on Gaussian Mixture Modelling of a multi-scale fissure enhancement filter. A groupwise fissure prior is also constructed to assist the lobe segmentation in regions of grossly incomplete fissures. Results demonstrate the performance of the fissure segmentation and the state-of-the art accuracy of the lobe segmentation (**Chapter 4**).

- Local disease and deformation distributions are presented to better quantify the spread of the disease and its effect on lung biomechanics. Through the use of manifold learning and manifold fusion, I show that the distributions outperform conventional metrics to predict lung function impairment. I also present the possibility of constructing potential trajectories of disease progression of COPD in the manifold space of the distributions (**Chapter 5**).

- A pipeline for the regional analysis of lung structure and deformation that is applied to the study of exacerbation susceptible COPD. The study analyses lung features from a global to a lobar level to investigate whether regional analysis of the lung adds important information to characterise an exacerbation susceptible lung from CT. Various new features that correlate with spirometry are also presented. (**Chapter 6**).

- The application of disease progression modelling to COPD. A novel clustering algorithm called SuStaIn (**S**ubtype and **S**tage **I**nference) was applied to the COPDGene cohort. This has demonstrated the utility of applying disease progression modelling

to progressive lung diseases. The model detected two main subgroups of COPD with various clusters associated to them, which differ by the disease progression trajectory. These subgroups correspond to previous clinical knowledge on the subject of COPD subtypes. I also demonstrate the utility of early diagnosis in COPD management and provide preliminary results in its application to studying exacerbations of COPD (**Chapter 7**).

1.3 Thesis outline

This thesis presents research on novel methods for automated analysis of lung CT scans with a focus on COPD. This research has exploited various analytical tools; leveraging methods from image processing to machine learning to quantify and investigate new clinical hypotheses related to COPD.

In Chapter 3, I present a novel algorithm that aims to classify patients by considering the distribution of various local features across the lung. This was applied to exacerbation susceptible COPD and tested on three cohorts of patients from COPDGene. In Chapter 4, I present a lobe segmentation algorithm that is based on the probabilistic segmentation of the fissures and the creation of a groupwise fissure prior. State of the art validation results are presented for both the fissure and lobe segmentation. Chapter 5 extends the work in Chapter 3 and presents the concept of local disease and deformation distributions to better model the spread of disease and its effect on lung biomechanics. In Chapter 6, the methods from Chapter 4 and 5 are exploited to study differences between exacerbation susceptible and non-susceptible patients and assess if regional measurements associate with exacerbation susceptibility. In Chapter 7, an event-based model called SuStaIn (**S**ubtype and **S**tage **I**nference) is applied to investigate different trajectories of COPD disease progression. In Chapter 8, I summarise and discuss the main results and contributions and present future directions for this project.

Chapter 2

Background

Within this chapter, I first introduce Chronic Obstructive Pulmonary Disease (COPD). The aim is to outline how it affects the lung, how it is currently diagnosed and the main treatment options. I also discuss the complexities faced in its diagnosis, which is motivating the development of various new tools such as those from medical image analysis. I then introduce how images of the lung can be acquired from Computed Tomography and Magnetic Resonance Imaging, which are the two main modalities used for COPD diagnosis and research. I then summarise the main efforts in COPD image analysis and present the main technological developments that are enabling researchers and clinicians to better study COPD. More focused discussions on the state of the art can be found in Chapters 3, 4, 5, 6 and 7.

2.1 Chronic Obstructive Pulmonary Disease

2.1.1 Pathophysiology of COPD

Chronic Obstructive Pulmonary Disease (COPD) is defined as the progressive development of airflow limitation that is not fully reversible. The main components of COPD are small airways disease and emphysema. COPD affects the respiratory system in a heterogeneous fashion, leading to variety of clinical symptoms with numerous extra-pulmonary comorbidities [158]. The relative contribution of each component varies widely across subjects, demonstrating an important subject-specific variability in the progression of disease [165].

Studies have shown that the small airways are the major sites of airflow limitation in COPD [97, 96], which correlate well with the pathological alterations in tissue composition. These can be attributed towards a long-term exposure of toxic gases and particles [165], with a strong historic association to tobacco smoking [12]. This leads to airway inflammation and mucus hyper-secretion, which is caused by a cascade of immune-mediated

events [96]. Tissue remodelling is a by-product of the immune system response. Within this process, damaged tissue is repaired by restoring the epithelium and the micro-vasculature whilst adding a connective-tissue matrix. This tends to thicken the airway wall by a progressive narrowing of the luminal area, which may conclude in a total occlusion of the airway [12]. This can lead to gas trapping, which is a pathological retention of air that cannot be expired [121]. The mechanical repercussion of inflammation and remodelling are key determinants in the pathophysiology of COPD. Laminar flow within the respiratory bronchioles is governed by Hagen-Poiseuille flow. As a result, airflow resistance varies according to the fourth power of the luminal radius. In a healthy state, the small airways are seen to exhibit low airflow resistance, with a total contribution of 10% with respect to total airway tree resistance [123]. However, narrowing of the small airway luminal area yields an increase in resistance, which is reflected by airflow limitation.

The second major component of COPD is emphysema. This is defined as the progressive destruction of the alveolar architecture. It is categorised pathologically as an abnormal or progressive enlargement of the distal airspaces (generation > 21) [6]. These lesions can be easily visualised in CT images as contiguous areas of attenuation close to the Hounsfield Unit (HU) of air (-1000). Emphysema risk factors are similar to those affecting the airways. These include cigarette smoke, environmental irritants, genetic factors and pollutants [12]. A variety of theories have been proposed to explain the principal mechanisms of emphysema [202]. These include the hypothesis that tissue destruction occurs as a result of an imbalance between protease and anti-protease activities or through oxidative stress and matrix remodelling [202]. These mechanisms lead to severe alterations in the load-bearing components of the lung parenchyma, which severely impacts the stress-bearing properties of the lung and leads to a loss of elasticity [201]. Moreover, tethering forces which may help airway distensibility are lost [50]. This causes airflow limitation [147] and impaired diffusing capacities [134], which are important determinants in COPD mortality.

2.1.2 Exacerbations of COPD

Exacerbations of COPD significantly add to the clinical and economic burden of COPD. They are a prime example of the heterogeneity of COPD. They are acute events that can dramatically alter the normal time-course of COPD [186]. They are a major determinant in the mortality, morbidity and quality of life of COPD patients [234]. Reducing the frequency of exacerbations and improving the efficacy of therapeutic interventions are key

future targets in the management of COPD.

A potential dichotomy in COPD exists based on susceptibility to exacerbations. Although exacerbation frequency is seen to increase with COPD severity, there is an independent susceptibility phenotype [101] associated with higher severity of disease and a prior history of exacerbations. Patients with a history of 2 or more exacerbations per year have a high likelihood of future exacerbations whilst those with less are more likely to experience no future episodes. Identification of the exacerbation susceptible subtype facilitates isolating a specific high-risk group to identify novel preventative strategies for personalised medicine.

2.1.3 Diagnosis of COPD

The Global Initiative for Chronic Obstructive Lung Disease (GOLD) have defined four GOLD stages [227], which represent a simple set of spirometric tests that define the severity of airflow limitation attributed to COPD (Table 2.1).

GOLD Stage	Severity	$^{FEV_1}/_{FVC}\%$	FEV_1%predicted
GOLD 1	Mild	< 70	≥ 80
GOLD 2	Moderate	< 70	< 80
GOLD 3	Severe	< 70	< 50
GOLD 4	Very severe	< 70	< 30

Table 2.1: Spirometric classification of COPD severity recommended by the Global initiative for Obstructive Lung Diseases (GOLD 2013) [227].

The spirometric indices important for the diagnosis of COPD are the Forced Vital Capacity (FVC) and the Forced Expiratory Volume in 1 second (FEV_1) (Figure 2.1). The FVC represents the maximum volume of air, which can be exhaled from a position of full inspiration. The FEV_1 measures the maximum amount of air exhaled in the first second of the FVC manoeuvre. Airflow limitation is defined in an individual when the ratio between FEV_1 and FVC ($^{FEV_1}/_{FVC}$) drops below 70%. Severity is assessed by considering the FEV_1, corrected according to reference values based on age, height, sex and race. This index is defined within this report as FEV_1%predicted. Recently, GOLD have introduced new guidelines [227] for the classification of COPD by adding a symptom-based feature to the spirometric indices. However, the old classification (Table 2.1) is used throughout this thesis.

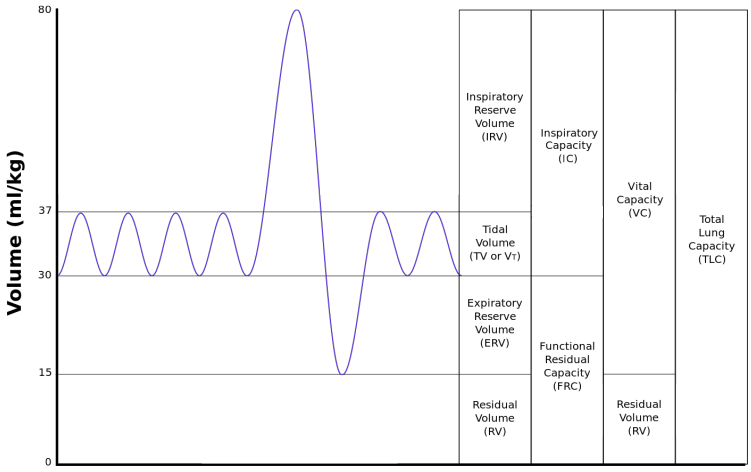


Figure 2.1: Flow volume versus time curve of breathing during a normal state and pulmonary function test state. Prior to the pulmonary function test, a patient is breathing freely, which corresponds to the resting tidal volume. The patient is then required to inhale maximally from tidal respiration to Total Lung Capacity (TLC) and then rapidly exhale in a forceful manner until no further volume is exhaled at residual volume (RV). The volume change from TLC to RV is the Forced Vital Capacity (FVC). The volume of air exhaled in one second is the FEV_1. Figure reproduced from Wikipedia Commons https://en.wikipedia.org/wiki/Spirometry#/media/File:Lungvolumes_Updated.png

2.1.4 Therapeutic practice of COPD

The management of COPD is largely reactive to pulmonary function tests. Assessment of disease severity and determining the need for treatment is based around the level of FEV_1%predicted. Whilst COPD is not curative, there are a various strategies that may be employed to slow down disease progression and relieve symptoms of COPD.

Smoking cessation is the most important intervention for modifying the course of disease [7] and efforts should be focused on helping patients to quit. In terms of a therapeutic intervention, bronchodilators are central to the management of COPD. Prevention and relief of symptoms can be achieved by inhalation of long-acting β agonists and can significantly improve symptoms, exercise capacity and health status in COPD patients [124] in addition to reducing exacerbation frequency [29]. Inhaled corticosteroids on the other hand have had a variable anti-inflammatory effect on patients with COPD with a lack of evidence to support that they modify the natural history of COPD. Surgical intervention is also possible through lung volume reduction surgery (LVRS). This practice aims to remove the least functional parts of the lung to improve lung function. It is most effective in patients with upper-lobe emphysema and has been shown to increase exercise capacity, reduced dynamic hyperinflation and improve pulmonary elastic recoil [124]. Effective patient stratification is

necessary for surgical planning since patients with a more homogeneous CT distribution of emphysema have a higher mortality with LRVS than conservative medical treatment [124].

2.1.5 The need for better diagnostic tools

Spirometry is the current gold standard for the staging of COPD severity and is also the most commonly used endpoint for drug development studies. There is a growing consensus that FEV_1 does not adequately describe the complexities of COPD [86]. FEV_1 is a global measurement and does not capture regional manifestations that may be implicated with disease progression [162]. It may also fail to characterise COPD heterogeneity, defined as the presence of both small airway disease and emphysema. Both phenotypes can occur synchronously and associate with airflow limitations [188].

It is also now apparent that various clusters of COPD exist, as demonstrated by unsupervised clustering [32, 27, 232]. Clusters of patients with similar levels of airflow obstruction can belong to different phenotypes with differing clinical characteristics as defined by symptoms, comorbidities and predicted mortality [27]. Genetic associations in COPD may be subtype dependent as seen in a cluster analysis using CT-features with a post-hoc analysis using genetic data [32]. These methods all used different patient cohorts, feature sets and clustering methods. However, the underlying theme of these studies is that a multi-dimensional assessment of COPD is needed upon which medical imaging may be a fundamental pillar. The current international guideline for therapy recommends adaptation of therapy based on severity of COPD, as assessed by post-bronchodilator FEV_1 and other symptoms [165]. This strategy may be correct for certain identified clusters but not in clusters with various extra-pulmonary comorbidities [27]. New end-points to evaluate therapy may be necessary and also better ways to stratify patients in clinical trials to assess the development of novel inhaled therapies.

2.2 Beyond spirometry: imaging lung structure and function

The major limitations of spirometry are significant arguments that the assessment of COPD should be multi-dimensional. The management of COPD, initiated with diagnosis and concluding with an accurate therapeutic strategy should rather follow the philosophy of P4 medicine (Personalised, Predictive, Preventive, Participatory) [2]. The use of imaging; through Computed Tomography and Magnetic Resonance Imaging is now emerging as a key determinant in precisely quantifying COPD. This is through the technological develop-

ments it is supporting such as segmentation, registration and pattern recognition that enable automated, robust and objective quantification of lung disease.

2.2.1 Computed Tomography

Computed Tomography (CT) is a diagnostic imaging procedure, which relies on X-ray emissions to acquire cross-sectional slices of the body (Figure 2.2). CT is based on the principle that tissue density can be measured by calculating the attenuation of a transmitted X-ray beam. The slices and attenuation values that are visualised in CT are reconstructed by measuring the transmission of X-rays through a series of detectors. The output is a three-dimensional representation of the underlying anatomy with respective macrostructural information relating to tissue density. Through various technological developments, it is now possible to acquire three-dimensional, high-resolution images with excellent contrast resolution that can be applied to various organs [114]. Despite the radiation-related cancer risks of computed tomography [16], it plays a major role as a tool for both diagnosis and research.

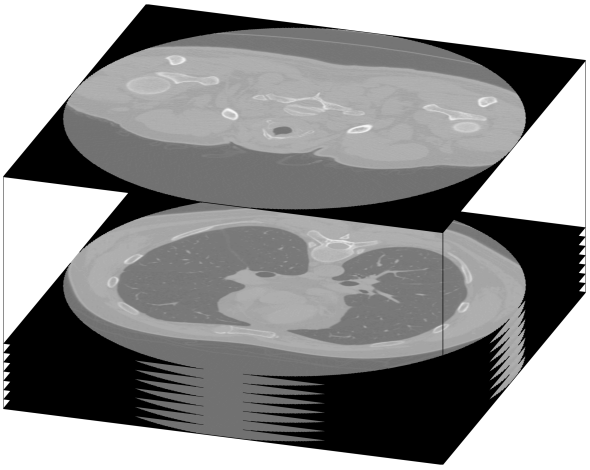


Figure 2.2: Three-dimensional CT image of the thorax created by stacking individual two-dimensional axial slices. Figure created using a control patient from the COPDGene study [166].

2.2.1.1 Fundamentals of Computed Tomography

The fundamental principle of CT is the differential absorption of X-rays by tissue [190]. To acquire images with sufficient signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR), there are three main requirements: 1) a sufficient number of X-rays need to be transmitted through the body for a high SNR, 2) X-ray absorption through different types of tissue needs to differ significantly to produce high CNR and 3) X-rays that are scattered

at random angles must be removed.

There are two main mechanisms that govern the behaviour of X-rays in tissue that affect the SNR and CNR of the image. They are the photoelectric attenuation and Compton scattering of X-rays. There are other mechanisms that occur such as pair production, Rayleigh scattering and photodisintegration. However, these can be ignored at the energy levels of conventional CT (0.1 to 1.5MeV).

Photoelectric attenuation describes absorption of the X-rays by the tissue. This provides contrast in the image since the probability of photoelectric interaction is strongly dependent on the effective atomic number of the tissue, the tissue density and the energy of the incident X-ray. Compton scattering refers to the interaction between the incident X-ray and a loosely bound electron in the outer shell of an atom. This collision causes the X-ray to scatter at a specific angle θ. The energy of each scattered X-ray photon is determined by a complex function linking incident photon energy and the angle of scatter θ. The reduction in beam intensity is proportional to the tissue electron density, which is approximately proportional to the physical density of the tissue.

Several factors in the imaging hardware affect the SNR, CNR and the spatial resolution of acquired images. However, photoelectric attenuation and Compton scattering are the principal determinants of X-ray attenuation and the grey-level of acquired images. The image stemming from CT is a parametric map of the linear tissue attenuation coefficient μ. Attenuation of N_0 incident X-rays through tissue can be determined by an exponential process that is a function of the distance travelled x such that

$$N = N_0 \exp\left(-\mu(E)\, x\right) \tag{2.1}$$

where E is the X-ray energy (keV) and N is the number of transmitted X-rays. The linear tissue attenuation coefficient is a function of the photoelectric absorption, Compton scattering and the energy of the X-ray:

$$\mu(E) = \mu(E)_{\text{photoelectric}} + \mu(E)_{\text{Compton}}\,. \tag{2.2}$$

The intensity of a voxel is directly related to the X-ray attenuation. The acquired image from CT is a parametric map of tissue attenuation. This map is obtained on a voxel-wise basis by image reconstruction. The reconstruction algorithm aims to determine the level

of attenuation in each voxel by analysing the fraction of X-ray beams ($\ln(N/N_0)$) that are measured in all detectors.

The intensity of a voxel in CT is a quantitative measure of the density of small discrete parts of the body. It is commonly expressed in Hounsfield units (HU). It is a dimensionless quantity, which is scaled based on the attenuation of air and water [57]. The radiodensity of air is usually given a value of -1000 HU whilst water is set to 0 HU. The attenuation value of a measured voxel is defined as

$$\mathrm{CT}_{voxel}(\mathrm{HU}) = 1000 \cdot (\mu_{voxel} - \mu_w) \,/\, (\mu_w - \mu_a) \tag{2.3}$$

where μ_w, μ_a and μ_{voxel} are respectively the attenuation values for water, air and the probed tissue. The attenuation values of water and air are generally calibrated by the manufacturer of the scanner. A general assumption is that the lung is composed of two materials: air and tissue/blood. A mixture model can be assumed, which states that the HU in lung CT images is a function of tissue and air content [95]. A linear relationship between HU and tissue density can thus be assumed although this breaks down in the presence of other materials such as fat and contrast agents. Radiographic values from CT consequently provide a non-invasive index of tissue density. This is particularly useful as it enables the segmentation and analysis of various pulmonary structures to quantify and study COPD.

2.2.1.2 Application to COPD image analysis

CT has great potential to influence the routine diagnosis and study of COPD. It has seen a rapid evolution in the quantitative evaluation of thoracic images [189, 222, 217]. It is the modality of choice for lung imaging as it permits direct evaluation of the pathological changes that contribute to airflow limitation. Novel developments in the segmentation of pulmonary structures [222], non-rigid registration [144], and pattern recognition with image classification [189, 217] are providing a rich array of tools for pulmonary image analysis.

2.2.1.2.1 Quantification of emphysema

The sensitivity of CT in probing tissue density has placed it as the optimal tool for quantifying emphysema. Various subtypes of emphysema exist (centrilobular, panlobular and paraseptal), which can be quantified by recognition of radiologic patterns in supervised [131] and unsupervised frameworks [131, 245]. However, emphysema is most frequently quantified by analysis of voxels with low attenuations. Emphysema is defined as the de-

struction of the lung tissue in the distal ends of the airway tree. This causes the voxel attenuation to decrease close to the value of air and appear as black regions in CT. Quantification of emphysema has typically been based on densitometric measures where statistical metrics stemming from the attenuation frequency distribution (Figure 2.3) are obtained. They have been used extensively in imaging studies of COPD, demonstrating strong associations with FEV_1 and GOLD severity [185], diffusing capacity of the lung for carbon monoxide (D_LCO) [79] and the 6-minute walk distance [49].

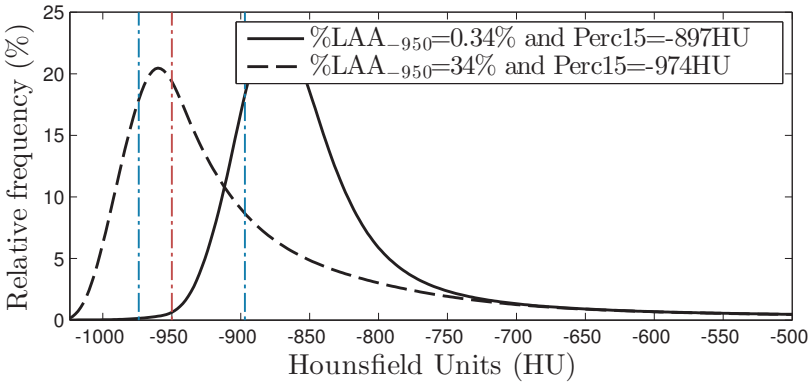


Figure 2.3: Frequency distribution of CT attenuation values with %LAA< -950 (red) and Perc15 (blue) thresholds. The distributions were obtained by fitting a cubic spline to a histogram obtained by quantising the CT attenuations in bins with a 25HU width. A threshold of -950HU is used to calculate the percentage of low attenuation areas (LAA), where percentage emphysema is defined as %LAA< -950HU. The Perc15 metric is the 15th percentile of the frequency distribution.

The most common approach relies on the application of a density mask [142]. This uses a threshold on the voxel intensity to calculate the percentage of low attenuation areas (%LAA). The most common threshold is -950HU as it has been shown to have the strongest correlations with microscopic evaluation and pulmonary function tests [74]. In addition to the %LAA, percentile densitometry is used, which seeks a HU threshold based on the percentile in the frequency distribution (Figure 2.3). The most common threshold is the 15th percentile based on studies of α_1-antitrypsin deficiency [156].

2.2.1.2.2 Quantification of small airway disease

The main contributor to airflow limitation in COPD are the small airways [96]. Their quantification is a major challenge in pulmonary image analysis and there is currently no gold standard for measuring small-airway disease. Accurate quantification of the smaller airways is inhibited by the resolution of modern CT scanners. Main developments in the assessment

of small-airway disease have thus focused on indirect measures. Remodelling of the airway wall leads to narrowing and eventually obliteration of the lumen of smaller airways. This will lead to an abnormal retention of air in the distal ends of the airway tree, which is known as gas trapping and is a surrogate marker of small airways disease.

Various techniques have been used to quantify the extent of gas trapping from CT. Gas trapping causes a decrease in the mean attenuation value of an expiration scan due to the retention of air after expiration [119]. A threshold of -900HU was initially used to quantify gas trapping in asthmatic patients [149]. A threshold of -860HU was then seen to have the strongest correlation with pulmonary function tests [126]. The threshold of -856HU (Figure 2.4) which is the current gold standard, was later adopted after a large-cohort study as part of COPDGene study [185]. The authors state that the mean attenuation of a normally inflated lung (6 mL air per gram of lung) is -856HU. Therefore, regions which are inflated above 6mL/g in a normal lung are morphologically abnormal. However, the origin of -856HU as a threshold remains unclear since there are no studies to my knowledge, which propose it as a threshold for gas trapping.

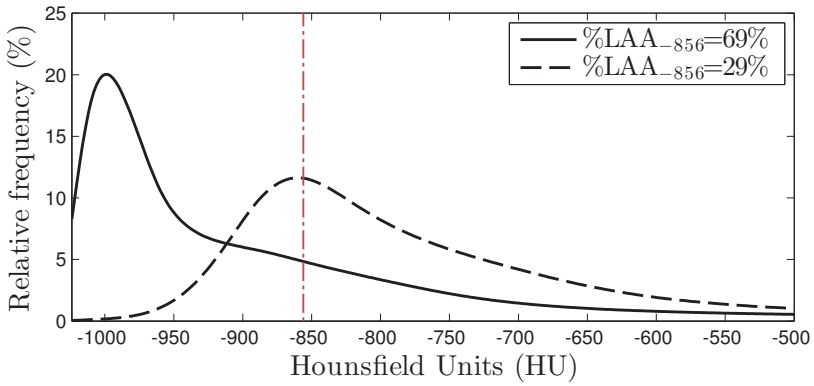


Figure 2.4: Frequency distribution of expiration CT attenuation values with %LAA< -856 threshold. The distributions were obtained by fitting a cubic spline to a histogram obtained by quantising the CT attenuations in bins with a 25HU width. A threshold of -856HU is used to calculate the percentage of low attenuation areas (LAA), where percentage gas trapping is defined as %LAA< -856HU.

Classic densitometric method do not compensate for the level of emphysema. They lack accuracy since voxels classified as gas trapping in expiration scans can be classified as emphysematous in corresponding inspiration scans. Various methods have been developed that exploit paired breath-hold scans such as the inspiratory-to-expiratory volume change of voxels with attenuations between -860 and -950HU (RVC$_{-860\,\text{to}\,-950}$) [126] and the ratio of mean lung attenuation from expiration to inspiration (E/I ratio) [59]. However, it is the

Parametric Response Mapping (PRM) technique [71] that has shown the greatest promise in quantifying small airways disease. It is based on the registration of paired breath-hold scans. The joint histogram of inspiration and expiration voxels yield a signature unique to the patient, which can be exploited to classify gas trapping that is not emphysema at inspiration (PRM_{fSAD}) and emphysema ($PRM_{emphysema}$). Whilst it has not yet been histologically validated, PRM measurements are reproducible over short periods of time [25], generally robust to scanner characteristics and misregistration [24] and have been shown to predict pulmonary function [163] and correlate longitudinal changes in FEV_1 [17].

2.2.1.2.3 Common limitations of tissue analysis

The attenuation value of a voxel is dependent on several factors such as radiation dose, scanner modality, the reconstruction algorithm and kernel, and the inspiration level [132]. This is an important issue specifically in longitudinal and multi-centre studies where controlling for all variables is difficult. Furthermore, the use of thresholds for quantifying emphysema (-950HU) and gas trapping (-856HU) in the lung may not be applicable. Various tools have emerged to mitigate these issues when faced with densitometry or the analysis of the voxel attenuations. These range from intensity modulation of the voxels [72] and volume correction of emphysema metrics [199] to data-driven modelling techniques for voxel-wise classification [84]. For instance, the lung attenuation signal (Figure 2.3) can be modelled as a mixture of skew-normal distributions, where the components correspond to emphysematous and healthy voxels [84]. This technique naturally builds on neuroimaging methods, which classify white matter, grey matter and cerebrospinal fluid (CSF) voxels by assuming a mixture of Gaussians [219]. This method [84] provides a more robust method for emphysema quantification since it removes the need for thresholding, which may not be suitable when faced with longitudinal scans or data stemming from different imaging protocols.

Data-driven techniques can also overcome further major limitations of classic densitometric techniques. The visual analysis of the lung parenchyma may suffer from intra and inter-observer variability [94]. This has motivated the development of techniques that may objectively quantify emphysema and its associated subtypes. This is primarily performed through the measurement of patterns of local features at the scale of secondary pulmonary lobule, which is called texture analysis. Within the context of pulmonary image analysis, texture analysis involves analysing lung intensity patterns in distinct regions of interest

(ROIs). Given appropriate labels that are obtained manually by a trained expert, a classifier can learn the appearance of various subtypes for future classification in new scans [131, 245, 194]. This may provide a more robust measure of lung disease that can differentiate structural differences between patients with equivalent lung function [243] or which may be complimentary to mean levels of emphysema [88].

A limitation of emphysema and small airways disease quantification is that they do not accurately quantify both the size distribution of disease clusters and their regional distribution in the lung. These can be important dimensions of COPD and may have important clinical implications. There have been attempts to quantify the size of emphysema clusters by analysing the size distribution of emphysema clusters [134] and it has been shown that a homogeneous presentation of emphysema can contribute to an accelerated decline in lung function independent of baseline FEV_1 [204]. In the context of small airway disease, there have been no reports on quantifying its appearance and distribution in the lung. Advanced analysis of small airways disease may help in our understanding of COPD progression such as at milder stages of COPD [17].

Lobar analysis of disease may also build on the limitation of global averages and can add important information on the disease state of the patient. It has been seen that lung volume reduction surgery may be particularly beneficial for patients with predominantly upper-lobe emphysema [66]. Regional analysis of emphysema has further shown that clinical manifestations of COPD may depend on the regional progression of emphysema [104, 75, 81, 180] however there are conflicting reports in the literature on the effects of upper-lobe and lower-lobe emphysema. However, it has been seen that a more homogeneous distribution of emphysema accelerates the decline of FEV_1 [204]. Despite necessitating the addition and quality control of lobe segmentation to the processing pipeline, regional analysis of COPD may prove to be crucial in advancing our understanding of COPD, as a staging tool for surgery and for the precise phenotypic characterisation of COPD from CT.

2.2.1.2.4 Quantification of pulmonary structures

To detect and quantify lung abnormalities from CT, it is necessary to first localise and segment structures of interest. Segmenting the lung is generally a prerequisite although there are structures outside of the lung cavity that can be analysed such as the pectoral muscles [89]. The high spatial resolution and CNR of modern multi-detector CT scanners permits

the visualisation of various pulmonary structures such as the airway tree, the vasculature and the lobar fissures (Figure 2.5). The segmentation of the airways, vasculature and fissures all have unique challenges and are active areas of development as illustrated by the EXACT09 [117], LOLA11 [99] and VESSEL12 [176] challenges.

Within this section, I focus on airway segmentation and its use in COPD analysis. An in-depth discussion on fissure and lobe segmentation can be found in Chapter 4 and lobar analysis of disease is discussed in Chapter 5. Whilst the segmentation of the vasculature can yield new insights on pulmonary vascular morphology [61], and is particularly useful in lobe segmentation, I focus on the challenges of airway segmentation and its clinical applicability since airway disease is one of the main components of COPD. An in-depth discussion of vessel segmentation techniques and applications can be found in a recent review by van Rikxoort and van Ginneken [222] and by Washko et al. [230].

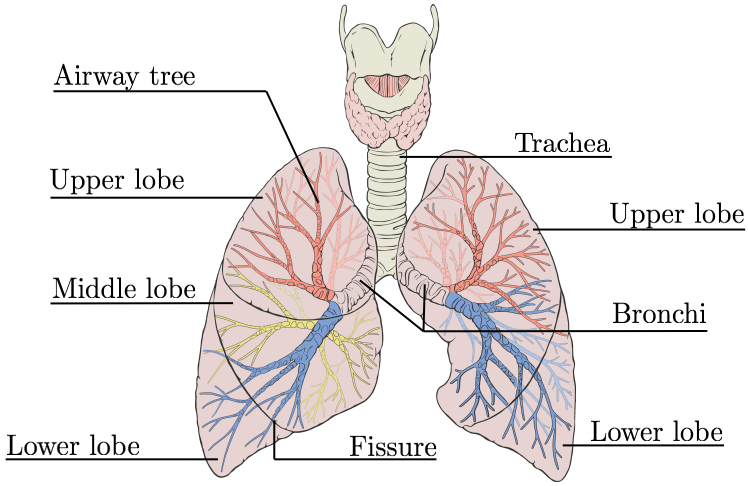


Figure 2.5: Illustration of the main constituents of respiratory system that can be segmented in CT. The right lung is divided into the lower, middle and upper lobes whilst the left lung is composed of the lower and upper lobes. During inspiration, air enters the trachea, which bifurcates into the main bronchi and further divides until the distal ends of the airway tree, where diffusion of O2 and CO2 occurs. Figure adapted from Wikipedia Commons https://en.wikipedia.org/wiki/Lung#/media/File:Lungs_diagram_detailed.svg.

Segmentation of the airways generally follows a two-step process. The airway lumen and the inner surface of the airway tree must first be identified. The airway tree in CT appears as a dark tubular, connected structure surrounded by bright walls. Methods generally exploit this property using grey-level region growing to iteratively grow the lumen from a seed point [218]. The airways become progressively smaller as they bifurcate, which can lead to partial volume effects. This can cause leaking of the airway tree and the segmen-

tation of spurious branches. Pathology such as collapsed airways and mucous deposition will terminate the region growing prematurely. A collapsed airway leads to a disconnected tree whilst mucous deposition and noise can lead to locally higher density, which inhibits the detection of distal branches. This had led to techniques that aim to prevent leaking through *explosion-control* [140], to detect and correct leaks through convolutional neural networks [38] or that are not based on intensity-based region growing such as supervised classification with a bronchi-enhancement filter [151].

The second step in airway segmentation exploits the segmented lumen to identify the outer airway wall surface. The most commonly used method is the full width at half maximum (FWHM). It is based on an analysis of the intensity profiles along a two-dimensional cross-section of the airway. It defines the wall boundary at the location where the intensity is half the peak value of the intensity profile. A limitation of this method is that it requires reconstruction of a two-dimensional cross-section of the lumen perpendicular to the airway centreline. Moreover, it can be biased by partial-volume effects and the reconstruction kernel [169]. Three-dimensional airway wall segmentation methods have therefore emerged, which rely on graph-cuts [161] and three-dimensional active surfaces [80].

Multiple airway measurements can be obtained from the segmentations that can be used to assess airway disease such as airway wall thickness, bronchial wall mean attenuation, wall area percentage (%WA) and internal perimeter (Pi). A standardised parameter called the Pi_{10} SRWA and Pi_{15} SRWA has been developed, which measures the square root of wall area (SRWA) for a hypothetical airway with an internal perimeter of 10 and 15mm [148]. Studies have shown increases in bronchial thickening in patients with COPD [178, 78]. Statistical modelling of airway tree structure and geometry has shown that airways from COPD and control patients stem from different distributions [65]. Increased wall thickness associated with FEV_1 [208] and other markers of airway abnormality correlates with functional markers such as the SGRQ and BODE index [125] and exacerbation frequency [87].

It is quite clear that airway segmentation has tremendous potential to monitor morphological changes attributed to COPD. Airway segmentations can be exploited for computational fluid dynamics (CFD) to better understand the effects of airway morphology on airflow [115] and assist in the development of aerosols used for drug deposition [211]. However, standardisation of measurement protocols is necessitated since studies often perform

measurements in different anatomical regions e.g. averaging across a generation versus measuring a single airway location.

2.2.1.2.5 Nonrigid registration and biomechanics from pulmonary CT

Nonrigid registration is an optimisation problem that aims to find the spatial transformation, which puts a floating image, into correspondence with a reference image. It is particularly suited to the diagnosis and characterisation of COPD. Longitudinal scans can be registered to a common space to locally monitor abnormal textural changes related to emphysema [77]. Biomechanical parameters that quantify lung deformation can be extracted from the registration of paired breath-hold scans [5] and may be predictive of patient outcomes in COPD [22].

Various registration algorithms focused on thoracic registration from CT have been developed. The EMPIRE10 challenge [144] and the DIR-Lab reference set [35] represent a good overview of the state of the art. Algorithms are now able to reach mean target registration errors[1] less than 1.0mm on paired breath-hold scans of both control and COPD patients. Novel developments in the literature are generally related to solving two major challenges: variations in intensity contrast caused by breathing [191, 247, 177, 92], and the sliding motion of the lung [93, 184, 155, 171]. Research into developing better optimisation strategies to find the minima of the optimisation are also ongoing [177, 93].

The deformation obtained from the registration can be exploited to investigate biomechanical properties of the lung, which may improve COPD characterisation from CT. Parameters that capture local volume change and heterogeneities can be derived such as the Jacobian determinant, the Anisotropy Deformation Index (ADI), [5], Jacobian heterogeneity [18] and measures from the Lagrangian strain tensor eigenvalues [107]. These biomechanical parameters have been shown to improve associations between emphysema and spirometry [18] but also associate with patients outcomes such as the SGRQ, 6-minute walking distance the BODE index [22].

The clinical utility of registration has also been shown when attempting to stage patients using CT-features in supervised classification [143, 23]. Addition of features only available through registration such as density change [143] or the ADI [23] significantly

[1]The target registration error (TRE) is the Euclidean distance between the coordinates of corresponding landmarks. A TRE equal to zero after registration means the transformation successfully put the landmarks in correspondence.

improve the performance of classifiers. This is not surprising as various studies have investigated significant associations between local volume change measurements and ventilation [168] or spirometry [43].

The application of registration to the study of COPD is not limited to the analysis of the deformation field. It is an instrumental process of the Parametric Response Mapping (PRM) technique, discussed in Section 2.2.1.2.2 that enables voxel-wise classification of emphysema and fSAD. Longitudinal scans may be registered to a common space to investigate the progression of emphysema [77, 198]. For instance, a local disease progression measure can be constructed to evaluate textural changes attributed to emphysema progression, which can track subtle local changes in tissue destruction. Moreover, registration is a cornerstone in the creation of statistical atlases. It may be desirable to construct an average model of lung function and structure such that normative ranges of structural and functional measures can be created [113]. This may provide an important basis for creating age-matched or disease-specific atlases to investigate how various covariates affect lung pathophysiology.

2.2.1.2.6 On the lack of longitudinal data to validate imaging research

A tremendous amount of effort has been made in the development of automated tools that aim to extract clinically useful data from Computed Tomography. This is evident in view of the methods presented within this chapter. Many techniques are aimed at extracting potential biomarkers that better predict lung function decline. For instance, Harmouche et al. [88] have presented a new emphysema metric based on the classification of various emphysema subtypes whilst the use of Parametric Response Mapping by Galbán et al. [71] has suggested that small-airway disease precedes emphysema in the progression of COPD. A significant limitation of these papers and by extension the work presented in this thesis is the lack of longitudinal data for validation. Most imaging studies are focused on cross-sectional data. Whilst this may provide a snap-shot of the disease process in a collection of individuals, it is not possible to extrapolate whether the findings better predict disease progression and whether they are suitable candidates as imaging biomarkers.

COPD is a progressive disorder that spans several decades. This has led to a lack of longitudinal data that permits effective validation of developed imaging biomarkers. A small minority of papers have performed analysis on longitudinal data at short time-scales (less than 3 years). These datasets are generally small such that there are questions on their

ability to capture heterogeneities present in the COPD population.

At the time of writing, there are several large studies that contain or aim to gather longitudinal data. These are: 1) ECLIPSE [224], 2) COPDGene [167] and 3) the Danish lung cancer CT screening trial [159]. Only ECLIPSE and COPDGene were focused on studying phenotypes and the progression of COPD. The ECLIPSE study is a 3-year longitudinal study that recruited 2,180 COPD subjects. It is not yet publically available but is intended to be submitted to the Genotypes and Phenotypes dbGaP repository of the National Institutes of Health. The COPDGene study is the largest multi-center study with over 10,000 patients recruited. However, the acquisition of data at 5-year follow-up has not been completed. In future work, the ECLIPSE and COPDGene studies present the exciting opportunity to be used for longitudinally studies of developed imaging biomarkers, the validation of disease progression models and the independent validation of research either performed on ECLIPSE or COPDGene.

2.2.2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) exploits the abundance of hydrogen protons (^{1}H) in water and fat to acquire images. Tissue contrast in a magnetic resonance (MR) scan is generally related to the proton density of various tissues. This differs from computed tomography, which relies on the ionising radiation of x-rays and the attenuating properties of various tissues to acquire images. It is therefore an exciting potential alternative and candidate for future routine diagnostic use in lung imaging. Due to the excellent soft tissue contrast of MRI, the lack of ionising radiation and its high spatial and temporal resolution, it is often the modality of choice when frequent scanning is necessary for diagnosis and therapy.

2.2.2.1 Image acquisition

Image acquisition in MRI is largely due to the interaction between the nuclear spin of a proton and the application of an external magnetic field [41]. A bulk magnetisation of the protons occurs when a strong, static magnetic field is applied to the body. It is the manipulation and detection of the proton magnetisation with a combination of radio-frequency pulses (RF) that permits non-invasive probing of the underlying macro and microstructure of tissue.

The intensity of a voxel is directly related to the proton density and magnetisation of the tissue. The cyclic application of RF pulses leads to a periodic proton magnetisation

and decay. The decay process is governed by several time constants (T_1, T_2 and T_2^*). These constants differ with tissue type and pathology, helping to produce contrast in an MR image. Acquisition parameters such as the echo time (T_E) and the relaxation time (T_R), which govern the application of the RF pulse and the measurement of signal decay are instrumental in setting tissue contrast.

There is a growing interest in the application of MRI for pulmonary research and COPD [20, 45, 62]. Due to the combination of morphological and functional information that MRI provides, it is a particularly attractive tool in the study of COPD. Probing the pulmonary architecture is however a challenging task. The proton density of hydrogen (^{1}H) is low in lung tissue because of the absence of both fat and water, which translates to a very poor ^{1}H signal using conventional MRI techniques. The alveolar network creates a multitude of air-tissue interfaces and causes significant magnetic field distortions (susceptibility artefacts), which weakens the strength of the acquired signal. Respiratory and cardiac motion complicate the acquisition process and further degrade the quality of pulmonary MR imaging. Whilst novel reconstruction methods [129] or protocols such as respiratory gating and rapid breath-hold imaging [239] attenuate these issues, the fundamental issue of proton density and susceptibility artefacts has inhibited the use of pulmonary MR imaging both as a diagnostic tool and a research-based method. Despite these issues, recent advances such as new acquisition protocols, hyperpolarised MRI and diffusion MRI are proving instrumental in pushing MRI as a key technique for the study of COPD.

2.2.2.2 Pulmonary proton MRI

The fundamental drawback of lung MR imaging relates to the apparent transverse relaxation time T_2^*. Local inhomogeneities in the magnetic field, caused by air-tissue interfaces at the alveoli cause a rapid dephasing of the signal. It can be as fast as 2ms in the lung at 1.5T [239] compared to averages of 35ms and 40ms for skeletal muscle and cartilage [197]. It may be appealing to increase the magnetisation from 1.5T to 3.0T to increase the signal-to-noise-ratio (SNR). Magnetic susceptibility is however sensitive to the magnetisation strength, leading to an even shorter T_2^* of 0.5ms [239]. Detection of the ^{1}H pulmonary signal has therefore only been historically achievable by including venous contrast agents to evaluate pulmonary blood flow and vessel hemodynamics [111].

Various methods have been designed that combine the relaxation signals inherent to various tissues and contrast agents to differentiate inflammation [228] from smooth mus-

cle remodelling and mucus deposition [111, 228]. Further paired studies between CT and contrast-enhanced MRI have shown the possibility of visualising gas trapping effects [13]. These methods are proving instrumental in quantifying the underlying pathologies associated with COPD although many more studies are needed for further validation.

The feasibility of automatically quantifying disease from pulmonary MR has been improved with the development of protocols such as ultra-short echo times (T_E) [127, 122]. This is facilitating the detection of traditional CT densitometric features such as emphysema [122] and diffuse lung disease [120]. Automated analysis is either performed by direct quantification of parametric T_2^* maps [152] or through direct assessment of the ^{1}H signal intensity [122]. It has also been shown that that COPD patients have significantly shorter longitudinal relaxation times [3] (T^1) and that parametric mapping of T^1 detects differences between emphysema and fibrosis [196].

To further tackle the challenge of low ^{1}H signal, methods applying optimisation techniques [103] and new protocols such as zero echo time (ZTE) [235] are producing images with similar qualities to those acquired with CT. More ingenious techniques have been proposed, which aim to reconstruct the image data and respiratory motion in a unified framework [129]. It has been hypothesised that relative changes in lung volume during tidal breathing causes relative changes in MR signal intensity [11]. This has been exploited in Fourier decomposition MRI (FDMRI) [47, 15], which disentangles the intensity changes due to the respiratory and cardiac cycles to obtain ventilation and perfusion-weighted images. Since the technique does not rely on the administration of contrast agents, FDMRI has strong potential for clinical translation.

2.2.2.3 Pulmonary hyperpolarised MRI

A major limitation in ^{1}H lung MRI is the challenge related to proton density and magnetic susceptibility. Whilst the development of ultra-short T_E, zero T_E and Fourier decomposition MRI are generating promising results, conventional ^{1}H lung MRI is at a disadvantage since it cannot directly image the terminal airspaces of the lung. This limitation can be overcome by the administration of hyperpolarised gases Helium-3 (^{3}He) and Xenon-129 (^{129}Xe) as contrast agents, which upon inhalation permits direct visualisation of lung airspaces [63]. Depending on the acquisition protocol and the nature of the quantitative evaluation, hyperpolarised MRI (HPMRI) facilitates direct quantification of not only pulmonary ventilation [106] but diffusion and other microstructural properties of the alveolar network [244].

Gas polarisation, typically through spin exchange optical pumping methods [179] causes a notable increase in nuclear polarisation and compensates for the low density of inhaled noble gas nuclei within the lung. This results in an increase in the MR signal, which can be detected with a scanner tuned to the appropriate resonant frequency. Ventilation images acquired via HPMRI are generally 1mm in-plane and $5-10$mm out of plane [46]. Helium-3 is currently the gold standard in hyperpolarised imaging since its polarisation is less technically challenging than in ^{129}Xe but also because it has a higher gyromagnetic ratio, which results in better signal quality. Global quantities of ^{3}He are limited and expensive, which is pushing the development of ^{129}Xe. In addition to the economic benefit of ^{129}Xe, it is also soluble in the pulmonary tissue barrier and red blood cell compartments, which makes it attractive as a tool for monitoring gas exchange and uptake in the presence of lung disease [56].

The most common application of HPMRI lies in quantitatively assessing ventilation maps of either ^{3}He or ^{129}Xe. In healthy adults, inspiration of ^{3}He or ^{129}Xe results in a homogeneous signal across the lung, which is evidence that all units of the lung are participating in the respiratory process. In contrast, areas of absent or low signal in the ventilation image correspond to *ventilation defects*, which define areas of the lung subject to destruction of the alveolar network and obstruction of the smaller airways [106]. This leads to a heterogeneous pattern of ventilation, which is a hallmark of COPD and can be used to monitor response to treatment [213].

The application of diffusion weighted MR pulses to hyperpolarised gases facilitates the quantitative evaluation of lung microstructure. Akin to the diffusion of water molecules in tissue, diffusion weighted HPMRI captures the random Brownian motion of gas atoms within the airspaces of the lung. At inspiration, gas diffuses within the lung microstructure. This can be captured and quantified with the Apparent Diffusion Coefficient (ADC) or extended to quantify morphological properties of the alveoli through microstructural-diffusion models [244].

The ADC is inversely related with the level of restricted diffusion in the lung. For instance, the ADC in a healthy lung is 0.2cm^2/s (^{3}He) and 0.04cm^2/s (^{129}Xe) compared to 0.86cm^2/s (^{3}He) and 0.14cm^2/s (^{129}Xe) in free unrestricted space [141]. In comparison, the ADC of an emphysematous lung is larger than a healthy lung, which reflects the enlargement of the distal ends of the airway tree. A large number of studies have shown the impressive

sensitivity of ADC to detect pathological alterations in lung tissue [241, 210, 40]. ADC in both ^{3}He [241] and ^{129}Xe [210] has strongly correlated with measurements from histology whilst differentiating between patients with emphysema and idiopathic pulmonary fibrosis. Importantly, it has been shown that diffusion-based markers can be useful in the pre-symptomatic detection of degraded pulmonary function [64]. This is of particular importance as it implies diffusion-weighted HPMRI could be potentially used as an early diagnostic test.

2.3 Summary

The use of medical imaging is now emerging as a critical tool to aid the routine diagnosis and management of COPD. The methodological tools that have arisen from both CT and MRI now facilitate going beyond FEV_1. Pulmonary imaging techniques such as HPMRI now provide a way to visualise and quantify regional improvements in ventilation over time and in response to therapy. However, a better understanding is necessitated between ventilation imaging and pulmonary structure-function abnormalities associated with disease progression.

Computed tomography and magnetic resonance imaging are the current best imaging modalities in view of studying pathological alterations of COPD and their effect of lung function. Both modalities have strengths and weaknesses in view of these aims. High resolution computed tomography has the ability to acquire high resolution scans at sub-millimeter voxel levels. Moreover, a plethora of methodological work has been performed on CT that enables a significant amount of information to be automatically extracted. However, the administration of ionising radiation limits its use especially in view of repeated follow-up scans and importantly in the peadiatric population. In contrast, the use of MRI in COPD imaging research is in its infancy. The main technological developments that arise in COPD-MRI research relates to image acquisition rather than image analysis. The spatial resolution of MRI is not yet at the level of CT. However, since it combines functional and morphological information and importantly does not adminster ionnising radiation, it is expected that MRI significantly challenges the use of CT as the dominant tool in COPD imaging research.

Exploiting the information that is inherent to expiratory and inspiratory CT scans now enables researchers to quantify more than just emphysema and gas trapping extent such as emphysema subtypes, an indirected measure of small airways disease, lung deformation

and measurements of the airway tree. A significant amount of information can thus be obtained through quantitative CT analysis. The main current challenge is to determine the features most sensitive to disease progression and which promote our understanding of the pathological changes that lead to impaired lung function. These features must be validated in independent studies, which is a difficult process as demonstrated by conflicting results based on upper and lower-lobe studies. This illustrates that better patient stratification for imaging-studies is imperative.

Future quantitative image analysis of the lung may be instrumental in the treatment of COPD. Despite the potential to improve our understanding of COPD, further methodological advancements are critically needed. There is a need to better capture the regional progression of disease, determine its effect of lung physiology and assess whether regional assessment of lung disease may help phenotype various subtypes of COPD. Moreover, pathological alterations in lung structure can differ locally. However, most clinical studies still only consider the mean extent of disease and ignore how local aspects of disease vary across the lung. By modelling these properties and extending the analysis towards lung biomechanics, there is a great potential to better phenotype COPD from CT, increase our understanding of the heterogeneities present in COPD and eventually improve therapy.

Chapter 3

Multiscale analysis of imaging features and exacerbation susceptible COPD

The work presented in this chapter is based on and reprinted, with permission from F. Bragman, J. McClelland, M. Modat, S. Ourselin, J. Hurst and D. Hawkes (2014)*. "Multi-scale Analysis of Imaging Features and its Use in the Study of COPD Exacerbation Susceptible Phenotype". In: *Medical Image Computing and Computer Assisted Interventions - MICCAI 2014.* Edited by P. Golland, N. Hata, C. Barrillot, J. Hornegger and R. Howe. Lecture Notes in Computer Science, vol 8675, pp. 417-424.

∗ **Contributions**

I conceived the method and developed the framework presented in the paper. I undertook and designed the experiments, analysed the data and wrote the first draft of the paper. Jamie McClelland and David Hawkes provided supervisory input and helped prepare the paper. John Hurst provided clinical input during the development of the methodology. Marc Modat and Sébastien Ourselin provided input on the use of NiftyReg for the lung registrations.

Abstract

In this chapter, I propose a novel framework for exploring patterns of respiratory pathophysiology from paired breath-hold CT scans. The aim is to quantify how imaging features are distributed throughout the lung in order to understand potential differences between COPD subtypes and perform image classification of these subtypes. Current classification algorithms seek various features that associate with pulmonary function tests and many exploit global values that ignore potentially important regional and local information. These techniques may be limited when attempting to classify patients that belong to different subtypes of COPD yet have similar levels of lung function impairment. The framework is therefore designed to address this limitation whilst enabling analysis of large datasets with the view of of determining relationships between CT features and disease state.

The algorithm is based on the local distribution of image features at various anatomical scales. These are sampled throughout the lung to build a distribution of local features. Principal Component Analysis is then used to visualise and quantify the multi-scale anatomical variation of features, whilst the distribution subspace can be exploited within a classification setting. This framework enables hypothesis testing related to the different phenotypes implicated in Chronic Obstructive Pulmonary Disease (COPD). I applied the framework to the study of exacerbation susceptible COPD. I tested whether patients who are exacerbation susceptible and non-susceptible can be classified at equal levels of disease severity.

3.1 Introduction

Exacerbations of Chronic Obstructive Pulmonary Disease (COPD) are defined as a sudden worsening of symptoms, which accelerate the decline in lung function leading to an increased risk of mortality. Understanding their pathophysiology is critical for predicting the patients at greatest risk of hospitalisation. Recent work suggests that the frequency of exacerbations is a distinct phenotype [101]. This is described as an exacerbation susceptible phenotype, where a patient may exhibit distinct physiological patterns resulting in an intrinsic susceptibility.

Recent studies have suggested a potential link between changes in lung structure, function and exacerbations. An association between emphysema progression and exacerbation susceptibility has been observed over a 2-year study [203] although none was observed in a larger 3-year study [44]. Changes in airway wall thickness, a primary determinant of airflow resistance has also been associated with exacerbation frequency [87] whilst pulmonary arterial enlargement relates to severe exacerbations [236]. Further, regional ventilation defects have been observed prior to acute exacerbations [105].

These studies have suggested a potential dependence between abnormalities in lung structure, the distribution of disease and exacerbations. This suggests that features from analysis of Computed Tomography (CT) scans may be extracted that may be sensitive to exacerbation frequency and may have potential in classifying patients that are exacerbation susceptible. However, traditional classification algorithms may not be suited to this problem since they are generally focused on automatically staging COPD severity, as defined by the GOLD stage. This has motivated the algorithm presented in this chapter. Various classification algorithms that have been applied to COPD are now discussed to put into context the work presented within this chapter.

Classification of COPD from CT scans has either been performed by quantifying global metrics for the lung or through the analysis of local regions of interest (ROI). Methods that train classifiers on global averages generally build a feature set by combining various features that may be informative of COPD severity. Bodduluri et al. [23] combined global values of density-based, texture-based and biomechanical features and showed that lung tissue biomechanics can be useful in characterising COPD. This was further echoed by Murphy et al. [143], who performed a similar experiment at a global and lobar level, concluding that information stemming from paired breath-hold CT scans is needed to stage

COPD from CT.

The limitation of classifiers that exploit global averages is that they cannot capture local textural changes that are products of COPD disease progression [134] and that may differ between patients at equal levels of severity. They ignore the anatomical presentation of disease, which may be an important dimension of COPD. Studies in the past have focused on classifying local textural patterns in ROIs using various features such as wavelet transforms [48], local intensity distributions [131] and local binary patterns [194]. The locally classified results can be fused together to obtain a singular measure of severity [194, 88]. However, these classifiers heavily depend on the creation of a training set that requires manual annotation by trained expert.

Different approaches consider lung CT classification as a weak labelling problem. In the classical setting; each ROI has an associated label for classification. If one considers a lung CT scan as an ensemble of analysed ROIs, a global label based on pulmonary function tests can be propagated to each ROI. The strategy then depends on the methodology for ROI analysis and how a global label can be learnt from the selection of ROIs. Sørensen et al. [193] proposed a framework that estimates a global posterior probability of COPD. This was performed by fusing individual ROI posterior probabilities obtained through a k-nearest neighbours (kNN) classification and a distance metric based on a summed histogram dissimilarity. This however disregards the fact only a subset of ROIs may be affected whilst ROIs with early signs of disease may be treated as normal for subjects not yet diagnosed. Moreover, it is likely not to work on the exacerbation classification problem since the level of emphysema in susceptible patients may be variable. The ROI limitation has motivated further classification tools [192, 42]. Sørensen et al. [192] classified CT scans by assuming that all ROIs contribute when an image is compared to others by using bipartite graph matching between sets of ROIs. Cheplygina et al. [42] exploited the Multiple Instance Learning (MIL) framework to classify patients where classification is based on distances between sets of ROIs.

I propose a new framework for the analysis of lung pathophysiology and the classification of lung CT scans (Figure 3.1). The main motivation of my work stems from the main limitations of classifier trained on global features. I hypothesised that the spatial distribution of disease is a discriminating factor in the presence of pathology. There should be a complex interplay between various features throughout the lung. This distribution is likely to vary

based on the subtype and severity of disease and can be captured using a set of ROIs measuring various discriminating features. My method is based on the measurement of image features representing the biomechanics and density of tissue, using a sliding box window at various anatomical scales. This is to deal with the bifurcating nature of the respiratory system. I apply it to the study of exacerbation susceptible and non-susceptible patients. The distributions measured at multiple scales are exploited to investigate differences between subtypes whilst attempting to classify susceptible and non-susceptible COPD patients.

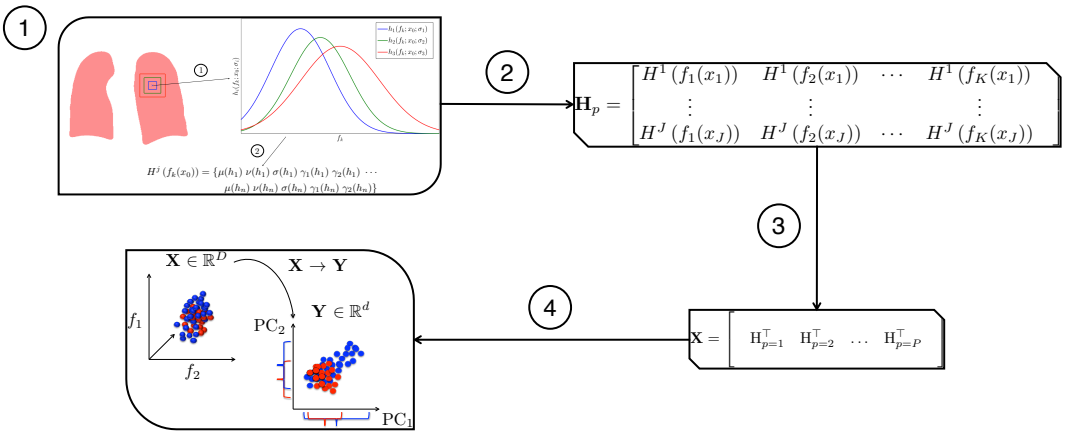


Figure 3.1: CT classification pipeline. 1) Features describing lung structure and lung biomechanics are sampled at various scales by computing the statistical moments of local distributions. 2) A matrix $\mathbf{H}_p$ for a patient p is created by analysing local distributions at J locations within the lung. 3) A population-wide matrix $\mathbf{X}$ is created by analysing P subjects. 4) Principal Component Analysis of $\mathbf{X}$ is performed to create $\mathbf{Y}$. The distribution of the principal component scores is used as a feature for classifying COPD patients.

3.2 Methods

3.2.1 Non-rigid registration

The NiftyReg registration platform [139] was employed to find the spatial mapping between the lung at full inhalation (Ω^*) and end exhalation (Ω). This was performed using a stationary velocity field, parameterised through a cubic B-spline interpolation. The Local Normalised Cross Correlation (LNCC) was the similarity measure whilst the bending energy of the velocity field was used as the regularisation. The registration was performed by considering only the lungs, delineated by segmented masks. The background volume was set to 0 Hounsfield Units upon which the masks were dilated to include a 0 HU border within the lung volume.

3.2.2 Feature extraction

The transformation $\boldsymbol{\varphi} : \Omega^* \rightarrow \Omega$, resulting from the registration serves to map each coordinate $\mathbf{x}^* \in \Omega^*$ to $\mathbf{x} \in \Omega$, such that the position of voxels at inspiration ($\mathbf{x}^* \in \Omega^*$) is known at expiration ($\mathbf{x} \in \Omega$). Biomechanical and density-based feature sets are derived using the information embedded within the transformation $\boldsymbol{\varphi}$, which maps all the voxels at expiration into their new position after deformation. The imaging features analysed within this chapter are summarised below in Table 3.1 and examples can be visualised in Figure 3.2.

Table 3.1: Feature set and measurements for lung CT analysis

Feature	Measurement
Biomechanical	Jacobian determinant (Equation 3.3, Figure 3.2c)
	Trace of Lagrangian strain eigenvalues (Equation 3.4)
	Variance of Lagrangian strain eigenvalues (Equation 3.4)
Density	Voxel intensity (HU) at inspiration (I_{ins}) (Figure 3.2b)
	Voxel intensity (HU) at expiration (I_{exp}) (Figure 3.2a)

(a) I_{exp} **(b)** I_{ins} **(c)** Jacobian determinant

Figure 3.2: Example feature maps. a) The lung at end expiration, b) at full inspiration and c) the Jacobian determinant that is obtained by registering I_{exp} to I_{ins} and analysing the transformation $\boldsymbol{\varphi}$.

3.2.2.1 Biomechanical features

To analyse the transformation $\boldsymbol{\varphi}$, I consider the deformation gradient tensor $\mathbf{F}$, which describes the deformation. It is defined as the gradient of the deformation in the reference frame Ω^*

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{x}^*} = \nabla_{\mathbf{x}^*} \boldsymbol{\varphi} \tag{3.1}$$

where at each voxel in Ω^*, it is defined in matrix format as

$$\mathbf{F}(\mathbf{x}^*) = \begin{bmatrix} \mathrm{F}_{11} & \mathrm{F}_{12} & \mathrm{F}_{13} \\ \mathrm{F}_{21} & \mathrm{F}_{22} & \mathrm{F}_{23} \\ \mathrm{F}_{31} & \mathrm{F}_{32} & \mathrm{F}_{33} \end{bmatrix} = \begin{bmatrix} \frac{\partial \varphi_1}{\partial x_1^*} & \frac{\partial \varphi_1}{\partial x_2^*} & \frac{\partial \varphi_1}{\partial x_3^*} \\ \frac{\partial \varphi_2}{\partial x_1^*} & \frac{\partial \varphi_2}{\partial x_2^*} & \frac{\partial \varphi_2}{\partial x_3^*} \\ \frac{\partial \varphi_3}{\partial x_1^*} & \frac{\partial \varphi_3}{\partial x_2^*} & \frac{\partial \varphi_3}{\partial x_3^*} \end{bmatrix} . \tag{3.2}$$

I derive 3 features from $\mathbf{F}$ to capture the respiratory process; the Jacobian determinant $\big(\det(\mathbf{F})\big)$ and the first 2 moments of the distribution of the eigenvalues of the Lagrangian strain tensor ($\mathbf{E}$). The Jacobian determinant is defined as

$$\det(\mathbf{F}) = \det\left(\frac{\partial \mathbf{x}}{\partial \mathbf{x}^*}\right) \equiv \frac{\Delta V}{V} \quad . \tag{3.3}$$

It measures the fractional volume change on a voxel-wise basis. The Lagrangian Strain Tensor $\mathbf{E}$ is derived from $\mathbf{F}$, by considering the Right Cauchy-Green Strain ($\mathbf{C}$)

$$\mathbf{C} = \mathbf{F}^\top \mathbf{F},\ \mathbf{F} = \mathbf{R}\,\mathbf{U},\ \mathbf{F}^\top \mathbf{F} = \mathbf{U}^\top \mathbf{R}^\top \mathbf{R}\,\mathbf{U} = \mathbf{U}^\top \mathbf{U} \ .$$

I am interested in analysing the stretches captured by $\mathbf{F}$. The tensor $\mathbf{C}$ results from a polar decomposition of $\mathbf{F}$, where the rotation component $\mathbf{R}$ is discarded by considering its orthogonal properties. The tensor $\mathbf{C}$ is thus rotation free, solely containing information about the stretches $\mathbf{U}$. The computation of the Lagrangian Strain Tensor ($\mathbf{E}$) follows

$$\mathbf{E} = \frac{1}{2}(\mathbf{C} - \mathbf{I}) \tag{3.4}$$

where $\mathbf{I}$ is an identity matrix. I derive the principal strains ($\lambda = \{\lambda_i \,|\, i = 1,2,3\}$) via an eigen-decomposition of $\mathbf{E}$. The trace ($\sum \lambda$), provides an overall measure of the magnitude of tissue strain whilst the variance ($\mathrm{Var}(\lambda)$) characterises anisotropy in the strain profile.

3.2.2.2 Density-based features

The transformation $\boldsymbol{\varphi}$ facilitates the analysis of the voxel densities (HU) at inspiration ($\mathrm{I}_{\mathrm{ins}}$) and expiration ($\mathrm{I}_{\mathrm{exp}}$) scans at corresponding locations. I consider the raw values of the intensities in $\mathrm{I}_{\mathrm{ins}}$ and $\mathrm{I}_{\mathrm{exp}}$ in addition to the percentage of emphysema ($\%\mathrm{LAA}_{\mathrm{ins}} - 950\mathrm{HU}$) and gas trapping ($\%\mathrm{LAA}_{\mathrm{exp}} - 856\mathrm{HU}$). In order to analyse the expiratory voxel intensities in the space of the inspiration scan, the transformation $\boldsymbol{\varphi}$ is exploited such that every voxel

in I_{ins} has corresponding expiration and inspiration intensities. Emphysema and gas trapping values are expressed as the percentage of voxels below -950 HU and -856 HU in a given volume. Within my framework, all values are analysed and calculated within local neighbourhoods across the lung, which is discussed in Section 3.2.3.

3.2.3 Multi-scale analysis of imaging features

I have a set measurements $f = \{f_k | k = 1, \cdots, K\}$ that define lung structure and function. Each measurement f_k (Table 3.1) provides information that is sensitive to COPD and lung structure. I have $K = 5$ image features: $f_1 =$ Jacobian determinant, $f_2 =$ eigenvalue trace, $f_3 =$ eigenvalue variance, $f_4 =$ inspiration voxel intensity and $f_5 =$ expiration voxel intensity.

I am interested in going beyond global averages that are tradtionally used as features in lung CT image classification. In this section, I propose a new methodology to sample each feature f_k throughout the lung to capture local feature distributions (Figure 3.3). These local distributions are parameterised to build a rich feature set that models local aspects of lung pathophysiology. The collection of local measurements stemming from each f_k represents the feature set that can be exploited as training data to build a classifier.

3.2.3.1 Feature distributions

I propose to sample the local variation of features (f_k) to quantify their distribution across the lung. This is performed by considering histograms ($h_i(f_k; x_j, \phi_i)$) of the local distributions of f_k. Each local feature distribution is centered at a voxel x_j ($j = 1 \cdots J$) within a neighbourhood ω governed by the scale ϕ_i, where $i = 1 \cdots n$ indexes the scale of the neighbourhood ω and j is the j^{th} sampled neighbourhood. Thus, distributions at increasing scales of analysis (ϕ_i) can be computed (Figure 3.3). The histograms are modelled by the first 4 statistical moments and the median. The feature f_k within ω centered at x_j is defined by:

$$\begin{aligned} H^j\left(f_k(x_j)\right) = \{&\mu(h_1)\ \nu(h_1)\ \sigma(h_1)\ \gamma_1(h_1)\ \gamma_2(h_1)\ \cdots \\ &\mu(h_n)\ \nu(h_n)\ \sigma(h_n)\ \gamma_1(h_n)\ \gamma_2(h_n)\} \end{aligned} \tag{3.5}$$

where μ is the mean, ν the median, σ the variance, γ_1 the skewness and γ_2 is the kurtosis. The vector $H^j\left(f_k(x_j)\right)$ thus contains the mean, median, variance, skewness and kurtosis of the local distribution of f_k centered at x_j and measured at the scales $\phi = \phi_1, \cdots, \phi_n$.

This process is repeated for all features $f = f_1, \cdots, f_K$ at locations $x = x_1, \cdots, x_J$ within

the lung. This enables to creation of a patient-specific matrix ($\mathbf{H}_p$) such that

$$\mathbf{H}_p = \begin{bmatrix} H^1\left(f_1(x_j)\right) & \cdots & H^1\left(f_k(x_j)\right) & \%\text{LAA}^1_{\text{ins/exp}}(x_j)\,\forall\phi \\ \vdots & & \vdots & \vdots \\ H^J\left(f_1(x_J)\right) & \cdots & H^J\left(f_k(x_J)\right) & \%\text{LAA}^J_{\text{ins/exp}}(x_J)\,\forall\phi \end{bmatrix}. \tag{3.6}$$

Each row of $\mathbf{H}_p$ represents the set of measurements for the feature set f at the sampled region centered at x_j. Each column of $\mathbf{H}_p$ represents the collection of a particular histogram measurement across the lung at sampled regions $x = x_j, \cdots, x_J$.

The Jacobian determinant (det($\mathbf{F}$)), the trace ($\sum\lambda$) and variance (Var(λ)) of the strain eigenvalues and the voxel densities in I_{ins} and Iexp are modelled locally across the lung ($k = 5$). I also incorporate the $\%\text{LAA} - 950\text{HU}$ and $\%\text{LAA} - 856\text{HU}$ for all ϕ_i, leading to $27n$ features per x_j. The number of sampled regions is determined by the sampling frequency of x_j at the finest scale (ϕ_1). For example, if there were $J = 1000$ sampled regions and $n = 3$ scales of analysis, $\mathbf{H}_p$ would be $\in \mathbb{R}^{1000x27\cdot3}$. It is the matrix $\mathbf{H}_p$ that is used as a feature within this chapter.

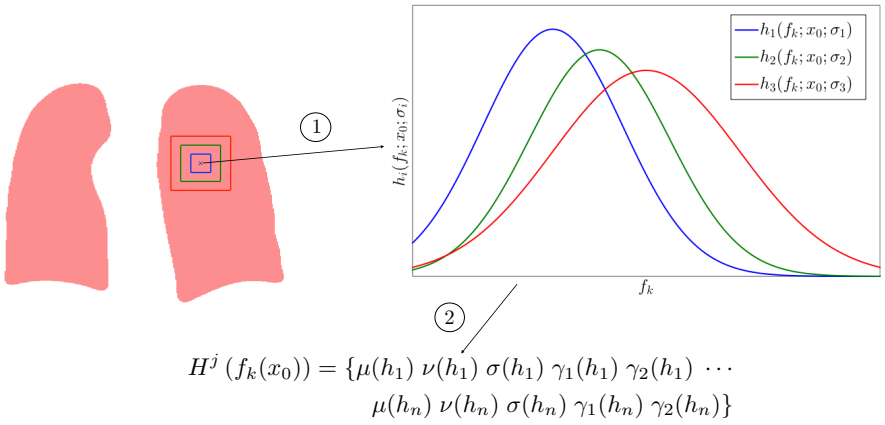


Figure 3.3: Illustration of the multi-scale sampling for a feature f_k. 1) A feature f_k (e.g. $\sum\lambda$) at x_j is sampled at $n = 3$ scales, leading to 3 local histograms $h_i(f_k;x_j,\phi_i)$. 2) Statistical moments and the median of $h_i(f_k;x_j,\phi_i)$ are calculated for all ϕ_i, leading to the set $H^j\left(f_k(x_j)\right)$. © 2014 Springer Nature. Reprinted, with permission from F. Bragman et al., Multi-scale Analysis of Imaging Features and its Use in the Study of COPD Exacerbation Susceptible Phenotype, Medical Image Computing and Computer Assisted Interventions, 2014.

3.2.3.2 Statistical feature analysis

A. Hypothesis testing using $\mathbf{H}_p$

Analysis of the distribution of values contained within each $\mathbf{H}_p$ allows hypotheses of changes in the global nature of local features to be made. For instance, consider the distribution of the variance of det($\mathbf{F}$) at all x_j. Each value demonstrates the local variation in volume change. The distribution of this measure across the lung will illustrate how the local variation is expressed, which may vary across subtypes. This facilitates a direct comparison of patient-specific distributions across phenotypes. To evaluate the distribution of various measures, one must analyse $\mathbf{H}_p$ column-wise.

B. Principal Component Analysis of $\mathbf{X}$

I am interested in modelling the distribution of parameters across the studied population. To perform this, the feature matrix $\mathbf{H}_p$ for each patient P is concatenated such that $\mathbf{X} = \begin{bmatrix} \mathbf{H}_1^\top & \cdots & \mathbf{H}_P^\top \end{bmatrix}$ where $\mathbf{X} \in \mathbb{R}^{N_J \text{ x } 27n}$ and N_J is total amount of regions sampled across all subjects P (Figure 3.4). I apply PCA on $\mathbf{X}$. This seeks a low-dimensional projection ($d << 27n$) of $\mathbf{X}$ into $\mathbf{Y}$ (Figure 3.5), where the variance of the projected features is maximised.

$$X = \begin{bmatrix} H^1(f_1(x_j)) & \cdots & H^1(f_k(x_j)) & \%\text{LAA}^1_{\text{ins/exp}}(x_j)\,\forall\phi \\ \vdots & & \vdots & \vdots \\ H^J(f_1(x_J)) & \cdots & H^J(f_k(x_J)) & \%\text{LAA}^J_{\text{ins/exp}}(x_J)\,\forall\phi \\ \vdots & & \vdots & \vdots \\ H^1(f_1(x_j)) & \cdots & H^1(f_k(x_j)) & \%\text{LAA}^1_{\text{ins/exp}}(x_j)\,\forall\phi \\ \vdots & & \vdots & \vdots \\ H^J(f_1(x_J)) & \cdots & H^J(f_k(x_J)) & \%\text{LAA}^J_{\text{ins/exp}}(x_J)\,\forall\phi \end{bmatrix}$$

Figure 3.4: Illustration of the population matrix $\mathbf{X}$. Each patient matrix $\mathbf{H}_p$ is concatenated to produce $\mathbf{X}$. The columns of $\mathbf{X}$ represent the different features obtained by analysis of the local distributions. The rows of $\mathbf{X}$ represent the local neighbourhood of a patient p. Each patient matrix $\mathbf{H}_p$ can be extracted from $\mathbf{X}$ by considering the row index i of $\mathbf{X}_{i,j}$. For the first patient in $\mathbf{X}$, $\mathbf{H}_{p=1}$ can be extracted by considering the sub-matrix $\mathbf{X}_{i=(1,\cdots,J),\forall j}$.

The entries of $\mathbf{X}$ are representative of the local histogram features measured at multiple scales. PCA of $\mathbf{X}$ allows me to compute the component scores within each neighbourhood defined by x_j (Figure 3.5). Thus, the computed scores can be projected to the image space to assess their distribution across the lung for each patient. Since the component scores are linear projections of the features measured at various scales, they will

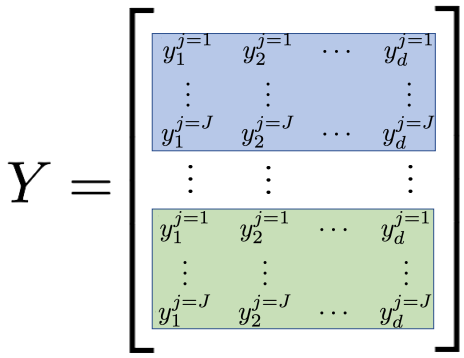


Figure 3.5: Dimensionality reduction of the population matrix $\mathbf{X}$ into $\mathbf{Y}$. The column-wise dimensionality of $\mathbf{X}$ is reduced from $27n$ to d by seeking a linear projection of the histogram features. The projection of each patient-specific matrix $\mathbf{H}_p$ within $\mathbf{X}$ into the new space of $\mathbf{Y}$ is known since the row indices of $\mathbf{Y}$ are equivalent to those in $\mathbf{X}$. Thus, the projection of $\mathbf{H}_{p=1}$ within $\mathbf{X}$ to $\mathbf{Y}$ can be found by $\mathbf{Y}_{i=(1,\cdots,J),\forall j}$. This allows one to project the principal component scores back into the image space since each y_i^j belongs to a region of the lung from a patient p. One can also calculate new features by considering the column-wise mean and variance of the linear projections.

capture potential fractal properties in line with the nature of the lung anatomy. The distribution of the principal component scores can be analysed to model patient-specific distributions by computing their respective mean and variance. This is done by calculating the column-wise mean and variance in $\mathbf{Y}$ for each patient (Figure 3.5). Thus, phenotype-specific distributions can be parameterised to produce a clinically meaningful classifier. Importantly, classification in the PCA subspace prevents overfitting as PCA removes colinearity in the features.

3.3 Experiments and results

The framework was tested on two sets of data to gauge its applicability in classifying exacerbation susceptible and non-susceptible patients. The first dataset was a pilot experiment ($P_1 = 20$), consisting of 10 patients in both groups with $f = 0$ and $f = 6$ exacerbations per year. The main motivation was to determine if differences in the local features could be detected between both cohorts and whether the distributions in the PCA subspace could be used as features to train a classifier. In the subsequent experiment, I increased the size of cohort to $P_2 = 49$ by including patients with a full spectrum of exacerbation frequencies ($f = 0, 1, 2, 3, 4, 6$). The aim was to determine if similar levels in classification performance could be achieved whilst better testing the performance of the classifier with repeated k-fold cross-validation.

3.3.1 Experiment 1: pilot study

3.3.1.1 Data

Inhale and exhale breath-hold CT images from the COPDGene study [167] were used. CT scans were acquired from multi-detector CT scanners, at full inspiration (200mAs) and at the end of normal expiration (50 mAs) with resolutions approximately equal to 0.66mm x 0.66mm x 0.73mm [167].

I tested the framework on $P = 20$ subjects with a GOLD 3 severity stage exhibiting $f = 0$ ($n = 10$) or $f = 6$ ($n = 10$) exacerbations per year. GOLD 3 patients were chosen due to their low variation in FEV_1. I chose two extreme sets ($f = 0$ and $f = 6$) to gauge the applicability of our framework in discriminating these phenotypes. The patient sets had a mean age of 60.2 and 67.5, a mean FEV_1%predicted of 42.1 and 40.5 and a mean $^{FEV_1}/_{FVC}$ ratio of 42.4 and 47.2.

3.3.1.2 Processing

Prior to the registration, the masks were dilated with a 3 voxel radius sphere. An analysis of the registration parameters was performed; demonstrating robustness in the registration to small parameter changes. This was performed by visual assessment of the results. A scoring system was not developed. A binary decision of pass or fail was used to determine whether varying the parameters such as control point spacing significantly affected the registration result. The assessment was performed by analysing the registration of key features such as the lung boundary, the vasculature, the airways and the lobar fissures. A registration failure generally occured when major errors at the fissure and lung boundary were detected. In some cases, a failure occured when there was a significant mismatch in the registered vasculature. A registration was deemed as a pass when only minor errors were present; notably at the level of the fissures and the vasculature. Minor differences in the registration quality were detected across registration parameters.

The standard deviation of the LNCC Gaussian kernel was set to 3^3 voxels, whilst the weighting of the regularisation was 0.05% of the overall optimised cost function. The finest control point spacing of the B-spline grid was set to 5 voxels along each axis. After registration, the inhale lung mask was eroded by a spherical element with a 7 voxel radius. This was performed to ignore regions prone to discontinuities and experience an extreme degree of motion, which may cause undesirable errors. The value of 7 was chosen by analysis of the Jacobian determinant maps. The majority of outliers could be removed by considering

an erosion with a spherical element of 7.

Sampling of the biomechanical and density features was performed using a cubic box window at scales $\phi_1 = 10$, $\phi_2 = 20$ and $\phi_3 = 30$ mm^3 ($n = 3$), which is consistent with the size of the secondary pulmonary lobule. A sampling frequency of 10mm in the x, y and z direction was used yielding approximately $7,500$ regions per lung. I ignored regions at all scales where 50% of the voxels fell outside the lung mask.

3.3.1.3 Hypothesis testing of the distribution of local features

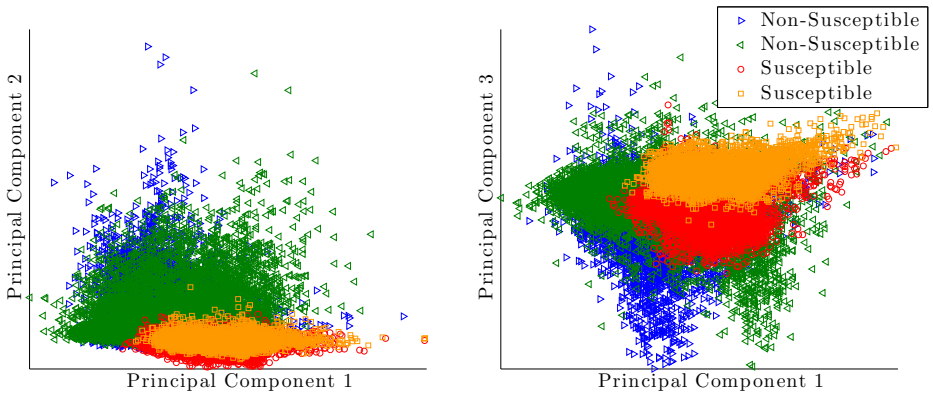
Feature distributions at the 3 scales ($\phi_1 = 10$mm, $\phi_2 = 20$mm and $\phi_3 = 30$mm) using $\mathbf{H}_p$ were investigated. I calculated the mean and standard deviation of each feature within $\mathbf{H}_p$ for all 3 scales. For example, the local mean inspiration intensity was sampled throughout the lung. Therefore, I calculated the mean and standard deviation of all sampled regions for each patient. This models the global distribution of local features across the lung.

As an exploratory analysis, I investigated if there were differences between these global measures of the locally sampled features across the susceptible and non-susceptible group. I performed a two-sample t-test for each subtype mean and standard deviation set to determine discriminating factors between both subtypes. I report the significant differences when $p < 0.05$.

A significant difference in the mean of $\sigma\left(\det(\mathbf{F})\right)$ ($.12 \pm .01$ and $.21 \pm .02$) at all scales of analysis was found ($p < .05$). The feature $\sigma\left(\det(\mathbf{F})\right)$ illustrates the variation in local volume change. The lower variation seen by the exacerbation susceptible group suggested that they exhibit a more homogeneous pattern in their volume change. No significance was seen in the standard deviation of $\sigma\left(\det(\mathbf{F})\right)$ ($p < .20$). I observed a marked difference ($p < .05$) in the mean ($.15 \pm .02$ and $.27 \pm .04$) of $\sigma\left(\sum(\lambda)\right)$ at all scales. This insinuated that for the susceptible group, the anisotropy in the magnitude of local tissue strain was more homogeneous compared to the non-susceptible patients. These suggested a possible distinction in deformation patterns, which were exploited in the classification. There were no significant differences in the density-based features.

Results from the PCA of matrix $\mathbf{X}$ corroborated the above, displaying evidence of distinct feature distributions across subtypes. (Figs. 3.6 and 3.7). Figure 3.6 illustrates 2 patient-specific principal component distributions for each subtype. These are characteristic of the phenotype distributions and are mostly consistent across each group. As the component scores are a linear projection of the features, Figure 3.6 suggests that there is

a consistent physiological pattern per subtype. This is illustrated by a variation in the heterogeneity of the scores as observed in the analysis of $\mathbf{H}_p$. This reinforces the notion of phenotype-specific distributions and the discriminating power of the distribution of disease.



(a) PC_1 versus PC_2 **(b)** PC_1 versus PC_3

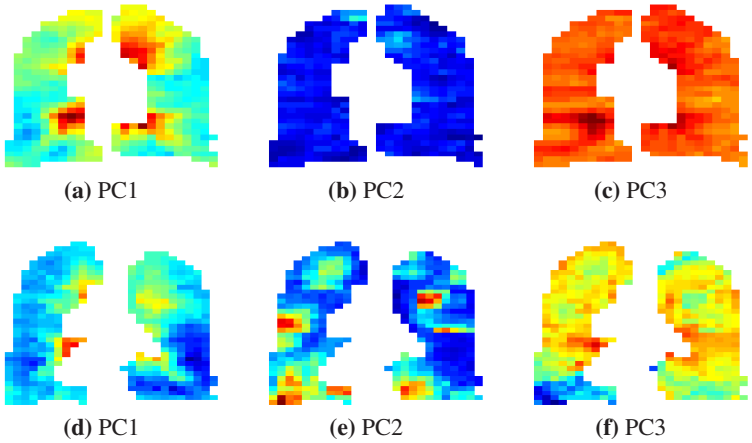
Figure 3.6: An example of the multi-scale principal component distributions for each sampled neighbourhood x_j for 4 patients (2 susceptible and 2 non-susceptible patients). The first 3 principal components explain $\approx 55\%$ of the variance of $\mathbf{X}$. © 2014 Springer Nature. Reprinted, with permission from F. Bragman et al., Multi-scale Analysis of Imaging Features and its Use in the Study of COPD Exacerbation Susceptible Phenotype, Medical Image Computing and Computer Assisted Interventions, 2014.

As the principal component scores were computed per sampled region (x_j) (Figure 3.4 and Figure 3.5), I was able to couple them with their respective anatomical location (Figure 3.7). This displays varying patterns in the physiology of the lung. As the principal components aimed to fully explain the lung macrostructure and the deformation captured within $\mathbf{H}_p$, these maps display a novel way of viewing how lung physiology differs between patients and across subtypes of COPD.

3.3.1.4 Classification of COPD exacerbation-susceptible patients

I aimed to classify exacerbation susceptible and non-susceptible patients based on the hypothesis that global and local patterns of disease differed across subtypes. This was highlighted in Figures 3.6 and 3.7, where a rise in feature homogeneity coincided with exacerbation susceptibility. I performed the classification on the feature projections using the mean and the variance of the principal component scores as features. The set explaining 90% (17/81) of the variance of $\mathbf{X}$ was chosen.

A leave-one-patient-out cross validation (LOPOCV) was employed to test the classifier. LOOCV iteratively selects one patient ($\mathbf{H}_{unseen}$) as the testing data whilst the remaining are used for training. I assumed independence amongst each training set during the LOOCV



(a) PC1 **(b)** PC2 **(c)** PC3

(d) PC1 **(e)** PC2 **(f)** PC3

Figure 3.7: Projection of the first 3 principal components of the multi-scale PCA into the image space. Coronal slice is at the mid-section. Top row: exacerbation susceptible phenotype. Bottom row: exacerbation non-susceptible phenotype. © 2014 Springer Nature. Reprinted, with permission from F. Bragman et al., Multi-scale Analysis of Imaging Features and its Use in the Study of COPD Exacerbation Susceptible Phenotype, Medical Image Computing and Computer Assisted Interventions, 2014.

to calculate accuracy and precision rates. For the classification, I projected $\mathbf{H}_{unseen}$ into the principal component space of $\mathbf{X}_{P-1}$ and used the mean and variance of the principal component scores as features. I used Support Vector Machines (SVM) as a classifier with a Gaussian radial basis function kernel $\sigma = 2.25$ and a soft-margin constant $C = 0.5$. Classifier performance from the LOOCV is the ratio of correctly classified labels to the total number of labels. Results from the LOOCV show that the framework has the ability to classify an unseen patient as either exacerbation susceptible ($f \geq 6$) or non-susceptible ($f = 0$) with a total accuracy of 75% (Table 3.2). After all iterations of the LOPOCV, 80% of the susceptible patients and 70% of the non-susceptible when *unseen* were correctly classified. This supports the applicability of the framework towards determining relationships between the distribution of disease with the clinical outcome.

Table 3.2: Classification results using Leave One-Patient-Out Cross Validation. © 2014 Springer Nature. Reprinted, with permission from F. Bragman et al., Multi-scale Analysis of Imaging Features and its Use in the Study of COPD Exacerbation Susceptible Phenotype, Medical Image Computing and Computer Assisted Interventions, 2014.

	Susceptible	Non-Susceptible	Total
Classification accuracy (%)	80	70	75

3.3.1.5 Summary

I have presented a novel framework for investigating global and local patterns of lung physiology in view of investigating pathological alterations in lung structure and deformation. The applicability of the framework in determining relationships between feature distributions and the severity of disease was shown through an analysis of the exacerbation susceptible and non-susceptible phenotypes. Analysis of the local feature distributions displayed significant differences in the nature of lung deformation across subtypes. This translated to subtype-specific distributions after dimensionality reduction, suggesting a potential intrinsic physiological behaviour attributed to both sets of patients. The main limitation of this pilot experiment relates to the validation since a relatively low number of patients was analysed. A larger population of patients is necessitated to correctly evaluate the performance of the classifier and demonstrate the clinical applicability of the framework.

3.3.2 Experiment 2: larger study

I subsequently tested the proposed framework on a new cohort of patients (Table 3.3) and a combination of patients used in the previous experiment with the new cohort (Table 3.4). I included patients with a whole spectrum of exacerbation frequencies ($f \in [0,2,3,4,6]$).

3.3.2.1 Data

I selected 29 additional patients at GOLD 3 severity from the COPDGene study (Table 3.3). I did not use all the data available for patients at the GOLD 3 level of severity. This is because it was desirable to have an unseen set of patients for use as validation in potential future follow-up experiments. The dataset within this new experiment can be classified as the training and testing set. At this level of severity, there were no patients exhibiting $f = 5$ exacerbations per year. There were also no additional patients with $f = 6$ exacerbations per year in the COPDGene cohort. Consequently, I randomly included 5 patients from experiment 1 to force a larger spectrum of susceptible patients. This led to 34 patients in dataset 1. The clinical characteristics and number of patients stratified by exacerbation frequency are shown in Table 3.3.

The second dataset consisted of all the patients analysed in experiment 1 (Section 3.3.1) in addition to the new patients in dataset 1 (Table 3.3) leading to a total of 49 patients. The clinical characteristics and total number of patients stratified by exacerbation frequency and susceptibility are shown in Table 3.4.

Table 3.3: Experiment 2: dataset 1

Frequency	Frequency Total	Total	Age	FEV_1%pred.	FEV_1/FVC
$f = 0$	16	16	64.51 (8.27)	38.61 (5.84)	45.94 (10.03)
$f = 2$	8	18	65.97 (8.07)	40.26 (6.15)	44.50 (7.74)
$f = 3$	2				
$f = 4$	3				
$f = 6$	5				
Total		34	$p = 0.62$	$p = 0.42$	$p = 0.64$

Table 3.4: Experiment 2: dataset 2

Frequency	Frequency Total	Total	Age	FEV_1%pred.	FEV_1/FVC
$f = 0$	26	26	65.68 (9.16)	38.66 (5.66)	44.88 (9.61)
$f = 2$	8	23	63.92 (8.68)	40.93 (5.78)	45.46 (8.71)
$f = 3$	2				
$f = 4$	3				
$f = 6$	10				
Total		49	$p = 0.48$	$p = 0.15$	$p = 0.82$

3.3.2.2 New registration pipeline

The original registration pipeline (Section 3.3.1.2) was tested on the additional patients selected (Table 3.3). After visual inspection, I determined that the registration algorithm was not robust on this larger cohort of patients. I observed various cases with unacceptable registration errors along the fissures and close to the diaphragm. Therefore, I investigated the original EMPIRE10 registration pipeline [136]. The EMPIRE10 pipeline consists of four registration stages (Figure 3.8). It is initiated with an affine registration [154] followed by three nonrigid registrations [138]. Each step is initialised by the previous registration.

The parameters for the three nonrigid registration steps that were investigated were: 1) the bending energy regularisation weighting, 2) the control point spacing at the finest resolution scale and 3) the LNCC kernel size. I investigated 150 different parameter combinations and visually assessed all results. No visual QA score was developed. The same process as outlined in Section 3.3.1.2 was performed. A simple binary rule (pass or fail) was produced for each registration result. Major registration errors at the lobar fissures and at the lung boundary represented failures. Minor errors in lung features such as at the airways and the vasculatures were ignored and the registration was classified as a pass. The registration parameters that produced the most consistent results across the cohort in Table

3.3 are shown in Table 3.5. I repeated the classification experiment in Section 3.3.1.4 to gauge the effect of the new registration pipeline on the initial cohort. I observed a 13% fall in the classification performance (Table 3.6).

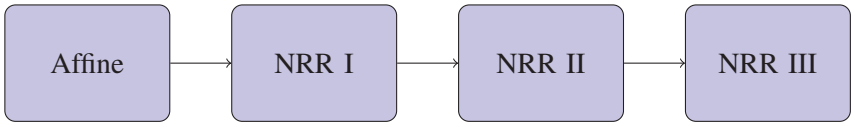


Figure 3.8: EMPIRE10 registration pipeline [136]

Table 3.5: Selected registration parameters.

	Energy weighting (%)	LNCC kernel (voxel)	Finest B-spline grid (voxel)
Stage 1	0.08	10	24
Stage 2	0.8	10	24
Stage 3	0.8	10	6

Table 3.6: Classification results using Leave One-Patient-Out Cross Validation on the initial cohort with the new registration pipeline (Table 3.5)

	Susceptible ($f = 6$)	Non-Susceptible ($f = 0$)	Average
Accuracy	0.70	0.60	0.65

3.3.2.3 Classification algorithms

I tested different classification models (Table 3.7). I considered linear or non-linear Support Vector Machines (SVM) whilst also performing no pre-processing and PCA pre-processing. A linear SVM requires training of a soft-margin parameter $\mathcal{C}$. This parameter is effectively a regularisation parameter in the SVM and acts as a trade-off between low errors on the training data and future generalisation of the model. It helps with cases when the data is not fully linearly separable and contains outliers. In a non-linear SVM, the data is first projected to a higher-dimensional space using a Kernel function. Linear classification is then performed in the new mapped space, which may help deal with datasets that are also not linearly separable. I again used a Gaussian kernel, which is parameterised by σ such that $\mathrm{K}(x_i, x_j) = \exp\left(-||x_i - x_j||^2/2\sigma^2\right)$ where x_i and x_j are instances of the training data.

3.3.2.4 Model training and validation

The methodology for training and estimating classifier performance is presented in Figure 3.9. The full dataset is denoted as $\mathcal{N}_T$. It is divided into training data ($\mathcal{N}_1$) and validation

Table 3.7: Classification models used in experiment 2

	PCA pre-processing	SVM
Model 1	✓	Linear
Model 2	✓	Non-linear
Model 3	×	Linear
Model 4	×	Non-linear

data ($\mathcal{N}_2$). A stratified split is employed such that the number of susceptible and non-susceptible patients in each set is approximately balanced.

A repeated k-fold cross validation (CV) is performed on $\mathcal{N}_1$ to train the model. The repeated stratified k-fold CV offers a simple platform to evaluate the model parameters, which generalise best over the data $\mathcal{N}_1$. In a stratified k-fold CV, the training data is randomly partitioned into k equal size subsamples $(f_1, \cdots, f_k)$ with an approximately equal proportion of group labels across folds. One sub-sample is retained as the validation set for testing the model whilst the remaining $k-1$ subsamples are used to train the classifier. This is performed k-times on all hold out sets. For the model training to be more rigorous, the k-fold cross validation is repeated n times, which leads to the creation of $n \times k$ hold-out sets. The set of parameters, which minimises the mean squared error across all $n \times k$ hold-out sets is chosen. Error in this experiment is defined as the complementary accuracy ($1-$ Accuracy). Accuracy is defined as the percentage of correctly classified labels in the hold-out set or $(TP+TN)/(TP+TN+FP+FN)$, where TP is the true positive rate, FP is the false positive rate, TN is the true negative rate and FN is the false negative rate.

Once the classifier has been trained and selected, it is applied to the unseen data contained within $\mathcal{N}_2$. This provides the quoted level of accuracy for the experiment. To avoid any potential bias when creating the datasets $\mathcal{N}_1$ and $\mathcal{N}_2$, the whole process is repeated Q times. This yields Q levels of accuracy. The final level of model performance is reported as $\mu(Q) \pm \sigma^2(Q)$ where μ is the mean and σ^2 is the standard deviation.

3.3.2.5 Results

All processing as described in Section 3.3.1.2 was applied to dataset 1 and 2. The registration pipeline presented in Section 3.3.2.2 was used. Datasets 1 and 2 were split into training and validation datasets with a 65% : 35% ratio. This was performed $Q = 5$ times. A $n = 5$ repeated $k = 5$-fold cross validation was performed to train the classifier in each dataset.

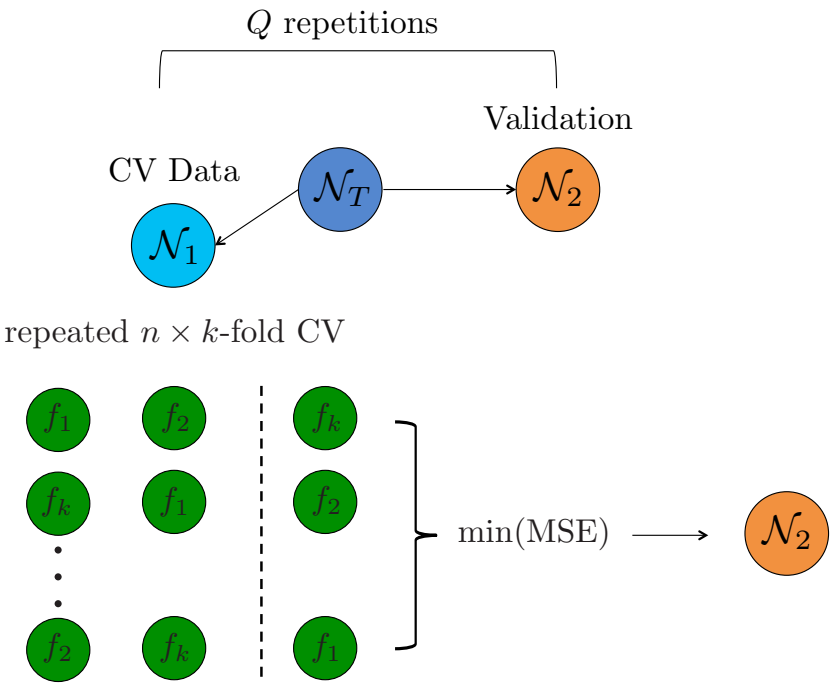


Figure 3.9: Repeated k-fold cross-validation procedure to train model and evaluate performance on validation set. The data ($\mathcal{N}_T$) is split into a training ($\mathcal{N}_1$) and a separate validation set ($\mathcal{N}_2$). Model parameter optimisation is performed using repeated k-fold cross validation setting. The classifier and the parameter choice that minimises the mean squared error on all hold-out sets over all repetitions is chosen. This classifier is then tested on the validation data ($\mathcal{N}_2$) to obtain a measure of accuracy on an independent set of patients. To avoid any bias in the creation of $\mathcal{N}_1$ and $\mathcal{N}_2$, the entire process is repeated Q-times.

This resulted in 25 testing sets to pick the model that minimised the mean-squared error. Within the cross-validation, I enforced an approximate 50% : 50% ratio between class labels in the hold-out sets ($f_1, \cdots, f_k$). The amount of patients used for model training and validation are shown below in Table 3.8. For the cross-validation, I chose the following range of hyper-parameters to train the SVM: $\mathcal{C} \in [0.05, 100]$ and $\sigma \in [0.5, 10]$.

Table 3.8: Number of patients in cross-validation and validation datasets.

Dataset	$\mathcal{N}_1$	$\mathcal{N}_2$
1	22	12
2	34	15

Results for dataset 1 and 2 can be viewed in Tables 3.9 and 3.10 respectively. The best level of performance in dataset 1 was 0.61 ± 0.08 (Model 3: No PCA, linear SVM). The results in dataset 1 displayed inconsistent levels of performance, which was reflected in fluctuating levels of accuracy across the Q repetitions. For instance, model 2 (PCA, non-

linear SVM) in dataset 1 exhibited excellent levels of accuracy (0.75) but also very poor levels of classification (0.17 and 0.33). I measured the value of the minimal mean-squared error in the cross-validation for each model and repetition. This is also displayed in Table 3.9.

The average level of performance across models in dataset 2 was equivalent (0.53 – 0.54). I also observed fluctuations in the level of accuracy across Q repetitions in all models. For instance, there was a large variation in the level of accuracy in model 4, ranging from 0.73 to 0.20. As seen in dataset 1, model 3 (No PCA, linear SVM) had the most stable performance with a standard deviation of (0.00 and 0.08). The value of the minimised mean-squared error for each iteration and model can also be viewed in Table 3.10.

Table 3.9: Classification results for dataset 1.

Model	Q_1	Q_2	Q_3	Q_4	Q_5	Mean CV_{error}	Average
1	0.58	0.58	0.33	0.58	0.58	0.27 ± 0.02	0.53 ± 0.11
2	0.75	0.42	0.75	0.33	0.17	0.31 ± 0.07	0.48 ± 0.26
3	0.58	0.58	0.58	0.58	0.75	0.26 ± 0.02	$\mathbf{0.61 \pm 0.08}$
4	0.58	0.58	0.33	0.58	0.43	0.26 ± 0.05	0.50 ± 0.12

Table 3.10: Classification results for dataset 2.

Model	Q_1	Q_2	Q_3	Q_4	Q_5	Mean CV_{error}	Average
1	0.53	0.60	0.40	0.60	0.53	0.19 ± 0.03	0.53 ± 0.08
2	0.80	0.33	0.60	0.33	0.60	0.27 ± 0.05	0.53 ± 0.08
3	0.53	0.53	0.53	0.53	0.53	0.19 ± 0.01	0.53 ± 0.08
4	0.20	0.73	0.73	0.60	0.40	0.29 ± 0.04	$\mathbf{0.54 \pm 0.23}$

3.3.2.6 Summary

The aim of this section was to determine the performance of the classification algorithm on a larger dataset. The experiment showed an expected drop in performance. In dataset 1, the best performance reached was 0.61 ± 0.08 across all repetitions. This was close to the baseline level of 0.65 achieved on the original dataset (Table 3.6). However, the performance of the classifier dropped further to 0.54 ± 0.23 when tested on the entire cohort (dataset 2). Whilst this level of accuracy is marginally better than random (accuracy = 0.5), the high variance of the classifier across the Q repetitions, combined with the generally stable

cross-validation mean squared error has suggested that the population is too heterogeneous, which may complicate training a classifier that generalises well. For instance, classification accuracies peaked at 0.75 and 0.73 in Tables 3.9 and 3.10 whilst dropping as low as 0.17.

A limitation of the pipeline, which in part may explain the performance of the algorithm relates to the feature construction. The mean and variance of either the multi-scale features or the principal components was used. This consequently assumes that the feature distributions are multivariate normal distributions. A more prudent approach would have been to either embed the classification in a dissimilarity space or employ the Multiple Instance Learning approach. For instance, if a dissimilarity-space approach is used, the classifier is trained in this dissimilarity space, which is built by considering pairwise distances between images. In the context of the proposed framework, the pairwise distance between two patients i and j would have been computed by considering distance between their respective feature matrices $\mathbf{H}_i$ and $\mathbf{H}_j$.

3.4 Analysis limitation

It has recently come to my attention that two of the biomechanical features used in this study are equivalent. The Jacobian determinant and the trace of the strain tensor eigenvalues both represent fractional volume change. This error was spotted when revisiting the mathematical formulation of the deformation gradient tensor and the Lagrangian strain tensor eigenvalues. The maximum principle strain should have been used, which is the largest strain tensor eigenvalue. However, this error does not change the validity of the results nor does it alter the theoretical foundations of the proposed framework. Excluding either of the features from the analysis is not expected to alter the chapter conclusions. For future studies, the biomechanical feature set should only include the Jacobian determinant (relative volume change) and the variance of the strain eigenvalues (deformation anisotropy).

3.5 Discussion and conclusion

Within this chapter, I have a presented a framework that quantified multi-scale distributions of features to attempt to classify exacerbation susceptible and non-susceptible patients. Analysis of features in local ROIs and exploiting their distribution throughout the lung has allowed me to capture potentially important information not attainable when using scalar values. Parameterising local density distributions encodes information about potential subtypes of emphysema [131] whilst the analysis of lung deformation at a local scale may

capture subtle effects of COPD on lung deformation.

Classification of COPD has traditionally been applied to predicting GOLD stage. To my knowledge, there were no examples in the literature aiming to classify exacerbation susceptible from CT scans. Traditional classification methods were not applicable in this case as they aim to quantify CT features that correlate with spirometry such as emphysema extent and mean lung deformation. This would not have worked on the chosen datasets since I analysed patients at equal stages of COPD classification. Whilst the results may seem poor considering the classification literature, this is the first time exacerbation susceptible COPD has been attempted to be classified. This work therefore sets the baseline level of performance for all ensuing work.

There are limitations in the study design that may have affected the obtained results. The classification problem tackled in this chapter was binary. This was motivated by the work of Hurst et al. [101] who defined the exacerbation susceptible phenotype. However, the absence of exacerbations in the past-year does not always equate to an absence of future episodes. Predicting exacerbation frequency using a Poisson or negative Binomial regression may have been preferential. An automated threshold then would have been applied to the prediction.

The classification results in both experiments also highlight the heterogeneity of the population. It has been observed that various COPD subtypes exist [27, 33]. Various CT-defined subtypes of COPD had higher likelihoods of exacerbation episodes despite different levels of emphysema and bronchial wall thickening [31]. Clustering of patients was not taken into account within this experiment. Therefore, the classifier was trained to recognise CT-features that predict susceptibility on a cohort of patients that may include various CT-defined subtypes. This would subsequently lead to a low cross-validation error on the training set yet high variable results on the testing set.

Future work should focus on analysing how the distributions vary with disease severity. The classification of exacerbation susceptible and non-susceptible patients is challenging and the difficulty is exacerbated when classifying patients at equal levels of severity. It is important to determine the clinical utility of the PCA multi-scale distributions. A future experiment would likely include a range of control patients in addition to those classified at GOLD stages 1, 2, 3 and 4. This would enable me to quantify a normative range of the multi-scale distributions in the control population and quantify how they vary with COPD severity.

This would provide the basis for understanding how sensitive they are to the clinical state of a patient.

Chapter 4

Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior

The work presented in this chapter is based on and reprinted, with permission from F. Bragman, J. McClelland, J. Jacob, J. Hurst and D. Hawkes (2017)*. "Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior". In: *IEEE Transactions on Medical Imaging*, vol. 36 (8), pp 1650-1663.

∗ **Contributions**

I conceived the method and developed the framework presented in the paper. I performed all the data processing. I designed and undertook all the experiments, analysed the data and wrote the paper. Jamie McClelland and David Hawkes provided supervisory input and helped in the preparation of the manuscript. Joseph Jacob performed the manual labelling to create the COPDGene reference set. John Hurst provided input on the clinical motivation of the algorithm in Section 4.1. Validation of the LOLA11 dataset was performed by the LOLA11 organisers.

Abstract

There is an important regional dimension in the anatomical presentation of COPD. The spread of tissue damage can evolve regionally and this may vary widely across the population. There may be discordant effects on lung physiology depending on whether regional damage in the lung is apical or basal. Moreover, lung deformation is expected to be heterogeneous across the lung since the level of deformation is expected to be higher close to the diaphragm. The framework developed in Chapter 3 was applied across the whole lung. However, it potentially ignored important region dependent information. To better characterise the severity of COPD from CT, it is necessary to perform the analysis on a regional basis. This has motivated me to develop a lobe segmentation algorithm.

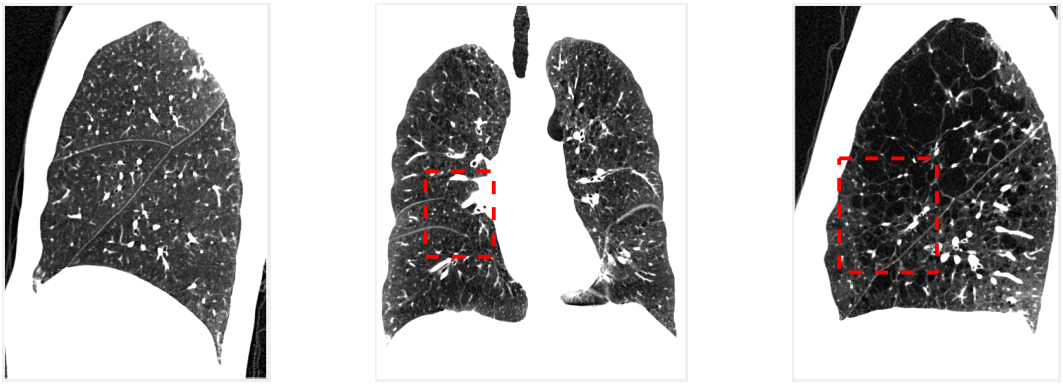
In this chapter, I present a fully automated, unsupervised lobe segmentation algorithm is presented based on a probabilistic segmentation of the fissures and the construction of a groupwise fissure prior. A two-class probabilistic segmentation segments the lung into candidate fissure voxels and the surrounding parenchyma. This is then combined with anatomical information and a groupwise fissure prior to drive non-parametric surface fitting to obtain the final segmentation. Validation of the fissure segmentation on a cohort of 30 patients from COPDGene is presented. The lobe segmentation was also validated on 55 cases from the LOLA11 dataset and 80 patients from COPDGene displaying state of the art performance. I also demonstrate the utility of including a groupwise fissure prior to help segment the lobes in regions of grossly incomplete fissures.

4.1 Introduction

Segmentation of the pulmonary lobes can facilitate the localisation and quantification of respiratory diseases and is of particular interest in Chronic Obstructive Pulmonary Disease (COPD). COPD can alter the structure of the lung through emphysematous destruction of lung parenchyma. The speed with which local pulmonary damage evolves can vary between patients with COPD [233, 206], yet the heterogeneity of local disease progression [233] may not be captured in lung physiologic indices that quantify function at a global level [206]. CT-based lobe segmentation provides an anatomically consistent reference frame for the quantitative analysis of parenchymal damage across large cohorts of patients and negates the requirement of a groupwise space for analysis. Knowledge of the underlying lobar distribution of disease may allow the identification of subtle COPD phenotypes or help identify patients that would benefit from interventions such as lung volume reduction surgery [36].

To characterise CT disease extent on a lobar basis, it is necessary to identify the pulmonary fissures. The fissures consist of invaginations of visceral pleura, which extend from the lung periphery to the lung hilum and separate the right and left lung into five lobes. The oblique and horizontal fissures divide the right lung into three lobes (upper, middle and lower) whilst the left oblique fissure divides the left lung into upper and lower lobes. When visible on CT, the fissures appear as bright, solid lines (Figure 4.1a). They represent two apposed layers of visceral pleura, which are usually devoid of airways and vascular structures [90]. However, the appearances of the fissures on CT can be variable in the general population [10]. A developmental failure of pleural invaginations can result in congenitally absent or incomplete fissures [90] (Figure 4.1b). Various pathological processes may damage the pleural surfaces disrupting the integrity of the fissures [90] (Figure 4.1c).

Automated lobar segmentation is most reliable when fissures are complete [53]. In cases with incomplete fissures, various methods have been developed that draw information from pulmonary anatomy and atlases. Lobe segmentation algorithms can be broadly categorised as either supervised [109, 118, 223, 220, 182, 174] or unsupervised [253, 54, 214, 255]. In our study, we extend the definition of supervised methods to encompass any algorithm that requires prior manual labelling to determine optimal fissure properties or the construction of anatomical atlases. Segmentation algorithms can be further subdivided on the basis of the segmentation of auxiliary structures. Methods can be



(a) (b) (c)

Figure 4.1: Example in the variation of fissure integrity. Scans with complete (a) and incomplete (b and c) fissures can be visualised. The incomplete fissures are due to fusion of lung tissue (b) or pathology (c). Slices are displayed in the intensity range $\mathcal{I} \in [-1024\,\text{HU}, -600\,\text{HU}]$. © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.

dependent [109, 54, 220, 253, 214, 255, 164] or independent [174, 150] of the information provided by the airway and vascular trees. Algorithms can also be classified based on their dependence on anatomical atlases[150, 220, 253, 174, 182] or whether the method is uniquely performed in the patient-space [109, 54].

Fissure segmentation can be classified as supervised when posed as a classification task or unsupervised when applied with a filter. A major shortcoming of filters is their reliance on arbitrary thresholds for segmentation. The inclusion of fixed segmentation thresholds may ignore potential fissure voxels or include excessive false positive voxels. Such thresholds are often compromised when lung attenuation values themselves are variably influenced by a range of factors including the CT reconstruction algorithm, CT slice thickness and patient inspiratory effort. The issue of the removal of false-positive voxels and its dependence on prior knowledge is a further limitation of several algorithms [109, 174, 238, 108]. The likelihood function of Lassen et al. [109] requires prior knowledge of fissure Hessian eigenvalues and may lead to an over-segmentation of fissure voxels that cannot be corrected through post-processing techniques. The work of Wiemker et al. [238] requires knowledge of the underlying Hounsfield intensity distribution of the fissures with no data-driven method presented to determine these parameters. Ross et al. [174] sample the image domain to detect the most likely fissure surface based on Hessian eigenvalues and a maximum a posteriori estimation. Their technique requires a lobe boundary shape model based on

manually segmented data to improve their fissure discrimination. The method is similar to the formulation of van Rikxoort et al. [222], which requires prior knowledge of manually labelled voxels to build a classifier. Manual annotation of data is time consuming and impractical in routine clinical practice. Moreover, it does not follow that a training dataset built on a single set of scans will generalise to a new cohort derived using different scanners, with varying reconstruction kernels. Such a constraint is also apparent in the likelihood function of Lassen et al. [109]. Our technique however learns the necessary model parameters from the volume being segmented, permitting the development of a robust segmentation tool, applicable across a broad range of datasets.

The use of prior knowledge derived from population models has increased in popularity [174, 150, 221, 254]. Zhang et al. [254] perform lobe segmentation using a single atlas search initialisation. The average fissure surface from a training set is exploited in a fuzzy reasoning system to segment the fissures and the lobes. An alternative multi-atlas selection mechanism has been proposed by van Rikxoort et al. [220]. This selects the most similar atlas to the patient by comparing the patient fissure segmentation to the atlas and exploits a transformation to combine atlas lobe labels with an approximate lobe segmentation. Ross et al. [174] exploit a deformable model in fissure surface extraction. The ability to exploit prior knowledge is an implicit advantage of atlases. However, if the training data is not large enough, this may not correctly model the shape variation within the population. These methods described all require complete segmentations prior to model building, which is a laborious task. I aimed to build a simple population model of the fissures negating the need for prior manual labelling without requiring complete fissure segmentations.

The limitations associated with the dependence on manually segmented data, either to train classifiers or build atlases was a major motivation of the work presented. When considering fissure segmentation, there is sufficient data within a single scan to detect the fissures when visible whilst rejecting most false-positives. In view of large-scale studies such as COPDGene [167], CT scans can also be pooled together to produce a prior, which negates the need for complete manual segmentations.

4.2 Methods

I present an automatic lobe segmentation algorithm (Figure 4.2) based on a probabilistic segmentation of the fissures (Section 4.2.2) and the construction of a groupwise fissure model (Section 4.2.3). Our study aimed to construct a fissure model (Section 4.2.3) using

complete and incomplete fissures to generate a confidence region based on a population. In the context of routine clinical care, new scans can be iteratively added if necessary to help strengthen the population model.

The main technical contributions of the algorithm are: a) unsupervised probabilistic segmentation of the fissures with iterative false-positive removal, b) the simultaneous construction of a groupwise prior without need for complete manual segmentations and c) post-processing of the airway segmentation to correct errors in seed labelling. An overview of the segmentation is shown in Figure. 4.2. The lungs, vessel and airway tree are first segmented. This is followed by a segmentation of the fissures using auxiliary tree structures as anatomical priors in a probabilistic setting. The segmented fissures are then combined using a groupwise registration framework to produce a population prior. The anatomical information, the segmented fissure and the groupwise fissure prior are then combined as a cost image for a watershed segmentation.

4.2.1 Data pre-processing

Lung masks are obtained with the algorithm of Hu et al. [100]. The vasculature is segmented by considering multi-scale vessel filtering [172]. The airways are segmented using region growing via evolution of a wavefront, which iteratively corrects for leakage across the airway wall [54]. It is assumed that the remaining structures after segmentation are the fissures and the parenchyma. A skeletonisation of the airways reveals the branching structure, used to label the lobar bronchi to generate surface fitting seeds. All pre-processing is performed using the Pulmonary Toolkit[1] with standard parameter settings.

4.2.2 Probabilistic fissure segmentation

I propose an unsupervised fissure segmentation framework that does not require any training data to classify fissure-voxels whilst negating the need to empirically determine algorithm parameters. I present a simple fissure enhancement filter that does not require any manual observations to set the parameters. I then construct a probabilistic framework to segment the fissures based on this enhancement filter. I assume a generative model between the observed filter result and the underlying segmentation of the lung and that these hidden segmentations exhibit separate Gaussian distributions. This is motivated by the work of Häme et al. [84] who first used mixture modelling with a hidden Markov Measure Field Model to segment normal parenchyma and emphysametous regions of the lung. The proposed model assumes

[1] https://github.com/tomdoel/pulmonarytoolkit

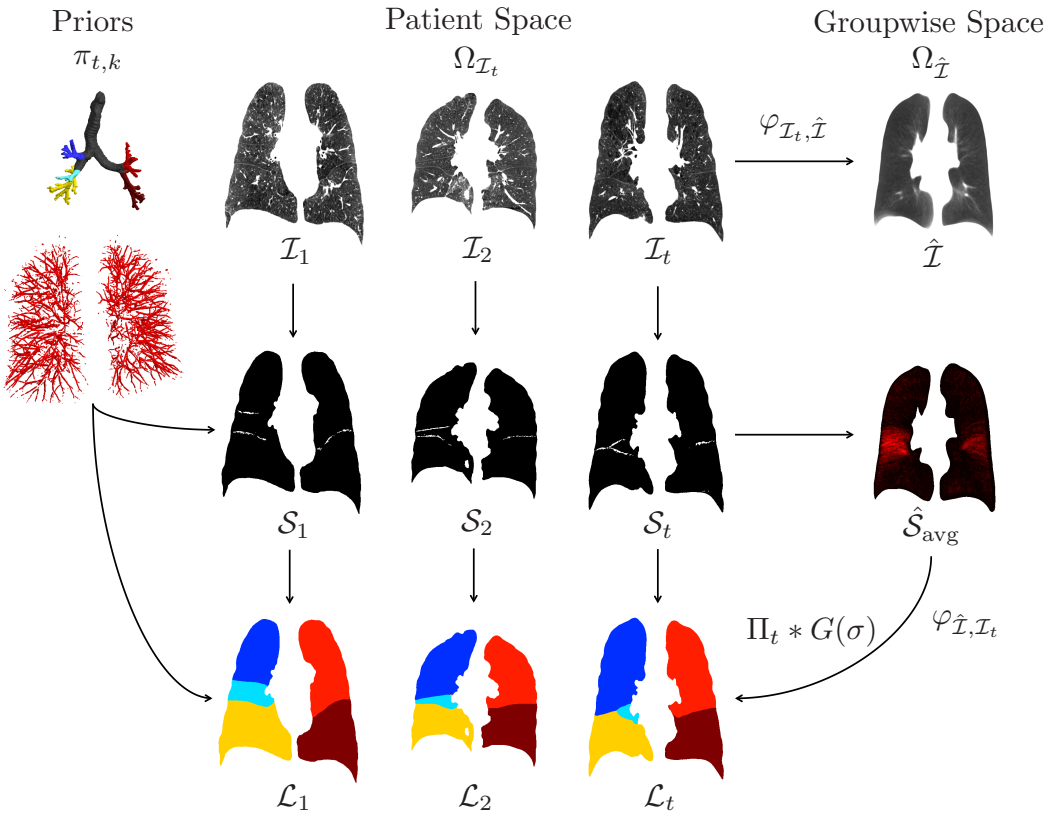


Figure 4.2: Lobe segmentation algorithm for processing a set of T patients, given by $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_t, \cdots, \mathcal{I}_T\}$. The lung mask, airway and vascular tree are segmented as preprocessing steps. Anatomical information ($\pi_{t,k}$, $k = (\text{fissure}, \text{tissue})$) is derived from the airway and vascular tree. A probabilistic segmentation of the fissures based on a filter ($\mathcal{F}_t$) exploits the anatomical information. This yields fissure segmentations ($\mathcal{S}_t$) for each patient t. For a given cohort ($\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_T\}$), a groupwise space is constructed yielding the set of transformations $\varphi_{\mathcal{I},\hat{\mathcal{I}}}$ to the common space $\hat{\mathcal{I}}$. This space is exploited to construct an average model of the fissures ($\hat{\mathcal{S}}_{\text{avg}}$). This is combined with the segmented fissures $\mathcal{S}_t$ and the patient-specific anatomical information ($\pi_{t,k}$) in a cost function for non-parametric surface fitting. Final lobe segmentations ($\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \cdots, \mathcal{L}_T\}$) are obtained for each patient (t) in the cohort ($\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_T\}$).

that the filtered image is generated by a two-class Gaussian mixture model (GMM), where the fissures and the parenchyma are the hidden segmentations that have generated the observed enhancement filter. Parameters of the GMM and the underlying segmentation are determined through application of the Expectation-Maximisation (EM) algorithm.

4.2.2.1 Multi-scale fissure enhancement filter

This filter aims to distinguish fissure-like voxels from surrounding structures. If one considers an image volume $\mathcal{I}_t$ from the set of images $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_T\}$, the Hessian matrix ($\mathbf{H}_t$) at a scale σ is obtained by considering the second derivative of $\mathcal{I}_t$ convolved with a Gaussian kernel $G(\sigma)$.

The width of the pleural cavity is likely to vary. The filter is embedded in a multi-scale framework to capture this variation. A voxel-wise eigen-analysis of $\mathbf{H}_t(\mathcal{I}_t;\sigma)$ yields scale-dependent eigenvalues λ_1, λ_2 and λ_3, which are, ordered based on their magnitude such that $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$, with respective eigenvectors $\hat{\mathbf{u}}_1$, $\hat{\mathbf{u}}_2$ and $\hat{\mathbf{u}}_3$. I adapted the vesselness filter of Frangi et al. [172] to capture voxels exhibiting a fissure-like shape.

As a sheet-like structure, a candidate fissure voxel will ideally be represented by a very large $|\lambda_3| >> 0$ with $\lambda_2 \approx \lambda_1 \approx 0$. The aim of the filter is to enhance voxels with this relationship whilst suppressing other auxiliary structures using the following parameters:

$$\mathcal{R}_a = \frac{|\lambda_1|}{\sqrt{|\lambda_2\lambda_3|}} \tag{4.1}$$

$$\mathcal{R}_b = \frac{|\lambda_2|}{|\lambda_3|}. \tag{4.2}$$

The parameter $\mathcal{R}_a$ helps differentiate spherical structures from plate-like and tubular structures. To differentiate plates from tubes, $\mathcal{R}_a$ is combined with the parameter $\mathcal{R}_b$, which seeks to enhance structures exhibiting a plate-like aspect ratio. A scale dependent filter (Eq. 4.3) is obtained by combining both terms:

$$\mathcal{F}_t(\sigma) = \mathbf{I}(\lambda_3(\sigma)) \cdot \left\{ \exp\left(-\frac{\mathcal{R}_a(\sigma)^2}{A}\right) \cdot \exp\left(-\frac{\mathcal{R}_b(\sigma)^2}{B}\right) \right\} \tag{4.3}$$

where $\mathbb{I}(\lambda_3(\sigma))$ is an indicator function such that $\mathbb{I}(\lambda_3(\sigma)) = 0$ when $\lambda_3(\sigma) > 0$ to seek only bright features. The parameters A and B control the sensitivity of each parameter $\mathcal{R}_{a,b}$. Small values of A and B (≈ 0) will only be sensitive to voxels with ideal values for the filter parameters ($\mathcal{R}_{a,b} \to 0$). Larger values for A and B ($\to 0.5$) will enhance voxels with less

ideal relationships with the caveat of enhancing more false positives (Section 4.4.1). A final value of $\mathcal{F}_t$ is found by computing the scale σ which maximises $\mathcal{F}_t$ at a voxel x:

$$\mathcal{F}_t(\mathrm{x}) = \max_{\sigma_{min} \leq \sigma \leq \sigma_{max}} \mathcal{F}(\mathrm{x};\sigma)\,. \tag{4.4}$$

There is a deviation from the ideal plate-like eigenvalue relationship in the presence of partial-volume effects, image noise and patient motion, which may result in a loss in the discriminating power of $\mathcal{F}_t$. We assume that both tissue classes are hidden segmentations that generate a range of values stemming from Gaussian distributions. I aim to capture these class distributions to accurately segment the fissures, even when the filter response is poor at a fissure voxel by considering local neighbourhood properties.

4.2.2.2 Fissure segmentation using a Gaussian Mixture Model

The output of the filter $\mathcal{F}_t$ is parameterised by a two-class GMM. The two underlying distributions in the signal correspond to the fissures and all other remaining structures. I assumed that the fissures and lung tissue are hidden segmentations (z) that give rise to the observed values y of $\mathcal{F}_t$. The segmentation can be modelled as a random process with a probability density function $f(z|\Phi_z)$ with parameters Φ_z. The total filter signal has a probability density function $f(y|z,\Phi_y)$ parameterised by the model parameters Φ_y. The goal is to estimate the segmentation z by the parameters $\Phi = \{\Phi_y, \Phi_z\}$. This is performed using the EM algorithm by estimating the maximum-likelihood parameters $\hat{\Phi}$ via maximisation of the log-likelihood

$$\hat{\Phi} = \arg\max_{\Phi} \log f(y|\Phi). \tag{4.5}$$

I considered the image model of Van Leemput et al. [219]. The index of a voxel x is $i \in \{1,2,...,n\}$ where n is the number of voxels within the lung mask. There are $K = 2$ classes (fissure and tissue). The class of the i^{th} voxel is defined as $z_i = e_k$. The variable e_k represents the class membership e.g. $e_{k=1}$ defines the fissure class and $e_{k=2}$ is the surrounding lung parenchyma. The response of the filter at voxel x_i is y_i. The filter values belonging to each class k are assumed to be normally distributed after log transformation with mean μ_k and standard deviation σ_k such that $\phi_k = \{\mu_k, \sigma_k\}$. The vector $\Phi_y = \{\phi_{k=1}, \phi_{k=2}\}$ represents the model parameters for both tissue classes. The overall probability density for y_i is defined as a mixture of normal distributions,

$$f\left(y_i \,|\, \Phi_y\right) = \sum_k \mathrm{G}_{\sigma_k}\left(y_i - \mu_k\right) f(z_i = e_k) \tag{4.6}$$

where G_{σ_k} represents the k^{th} class zero-mean normal distribution with standard deviation σ_k and $f(z_i = e_k)$ is the class prior probability of a voxel x_i. By assuming statistical independence over all voxels $x \in \mathcal{I}$, the overall joint probability density is given by

$$f(y \,|\, \Phi_y) = \prod_i f(y_i \,|\, \phi_y). \tag{4.7}$$

The maximum-likelihood estimates for Φ_y are found using Eq. 4.5 by seeking the parameters that maximise Eq. 4.7, giving the following update equations for the model parameters

$$\mu_k^{(m+1)} = \frac{\sum_i p_{ik}^{(m+1)} y_i}{\sum_i p_{ik}^{(m+1)}} \tag{4.8}$$

$$\left(\sigma_k^{(m+1)}\right)^2 = \frac{\sum_i p_{ik}^{(m+1)} \left(y_i - \mu_k^{(m+1)}\right)^2}{\sum_i p_{ik}^{(m+1)}} \tag{4.9}$$

where

$$p_{ik}^{(m+1)} = \frac{f\left(y_i \,|\, z_i = e_k, \Phi_y^{(m)}\right) f\left(z_i = e_k\right)}{\sum_{j=1}^{K} f\left(y_i \,|\, z_i = e_j, \Phi_y^{(m)}\right) f\left(z_i = e_j\right)} \tag{4.10}$$

is a probabilistic estimation of the hidden data z_i of class k at voxel x_i given the filter value y_i. The class k is iterated with the class index j and m denotes the EM iteration number.

The segmentation resulting from Eq. 4.10 can be sensitive to noise, image artefacts and false positives as the segmentation is only based on y. Priors ($\pi_{t,ik}$) that incorporate probabilistic information about the segmentation are typically added to the model [219, 28]. In the context of this work, the likelihood of the fissure location (Figure 4.4b) can be quantified based on the vessel and airway tree. I computed the vessel density (v_d) [54], which is obtained by applying a strong (10mm isotropic) Gaussian filter to the vesselness filter of Frangi et al. [172]. This measure is inverted and scaled in the range $[0,1]$ using min-max scaling such that regions of low vessel density are close to 1. Airway density (a_d) is estimated by computing the Euclidean distance transform to the airway segmentation and is normalised using min-max scaling to the range $[0,1]$ such that regions of high distance

to the airways are close to 1. The fissure likelihood measure is defined as $\pi_{t,ik=1} = {}^{(a_d+v_d)}/_2$ and the tissue likelihood is $\pi_{t,ik=2} = 1 - \pi_{t,ik=1}$. The subscript t is dropped in Eq. 4.11 and 4.12 for convenience. The anatomical information is integrated into Eq. 4.6 by setting $f(z_i = e_k) = \pi_{ik}$.

Information about lung structure, spatial smoothness and morphology can be also be enforced by considering a Markov Random Field (MRF) regularisation term (U_{MRF}). The probability of a voxel i belonging to tissue class k is now dependent on the first-order neighbours $\mathcal{N}_i$. The neighbourhood system at a voxel x_i is defined as $\mathcal{N}_i = \{\mathcal{N}_i^x, \mathcal{N}_i^y, \mathcal{N}_i^z\}$ in the face-connected neighbourhood. The likelihood term (π) is now augmented with an MRF that is dependent on the probability and curvature of neighbouring voxels. By employing the formulation of Van Leemput et al. [219], Eq. 4.10 is updated to

$$p_{ik}^{(m+1)} = \frac{f\left(y_i \,|\, z_i = e_k, \Phi_y^{(m)}\right) f\left(z_i = e_k \,|\, p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)}\right)}{\sum_{j=1}^{K} f\left(y_i \,|\, z_i = e_j, \Phi_y^{(m)}\right) f\left(z_i = e_j \,|\, p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)}\right)} \tag{4.11}$$

with

$$f\left(z_i = e_k \,|\, p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(}m)\right) = \frac{\pi_{ik} e^{-\beta_i U_{\text{MRF}}\left(e_k \,|\, p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)}\right)}}{\sum_{j=1}^{K} \pi_{ij} e^{-\beta_i U_{\text{MRF}}\left(e_j \,|\, p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)}\right)}}. \tag{4.12}$$

where the MRF term $U_{MRF}\left(z_i \,|\, p_{\mathcal{N}_i}, \Phi_z\right)$ is an energy function dependent on $\Phi_z = \{\mathbf{G}\}$ and the MRF weight β_i is kept constant for all voxels i.

Structure in the segmentation is enforced by considering neighbourhood probabilities ($p_{\mathcal{N}}$) and constraints on the curvature of the surface (f_{dot_k}) (Figure 4.3b). The fissure surface exhibits low local curvature, which will be captured in the eigenvector $\hat{\mathbf{u}}_3$. A weight in the MRF energy term is introduced based on the dot product of the eigenvectors of a neighbouring voxel x_l with the center voxel x_i such that

$$f_{dot}(x_l, x_i)_k = \begin{cases} 1 - \exp\left(-\frac{\left(|\hat{\mathbf{u}}_{x_l,3} \cdot \hat{\mathbf{u}}_{x_i,3}|\right)^6}{0.25}\right) & \text{if } k = 1 \\ \exp\left(-\frac{\left(|\hat{\mathbf{u}}_{x_l,3} \cdot \hat{\mathbf{u}}_{x_i,3}|\right)^6}{0.25}\right) & \text{if } k = 2 \end{cases} \tag{4.13}$$

When considering the fissure class ($k = 1$), the weight will tend to 0 as dissimilarity in the local curvature increases ($1 - \exp\left(-|\hat{\mathbf{u}}_{l,3} \cdot \hat{\mathbf{u}}_{i,3}|\right) \to 0$). If neighbouring voxels x_l and x_i have similar curvature, $f_{dot}(x_l, x_i)_{k=1}$ will tend to 1. The goal of this function is to force

candidate fissure voxels to have approximately equal curvature whilst negatively weighting the probabilities of false-positives with non-equal local curvature. The shape of the function within Equation 4.13 was chosen to have a highly non-linear form such that only voxels with $\left(|\hat{\mathbf{u}}_{x_l,3} \cdot \hat{\mathbf{u}}_{x_i,3}|\right) > 0.90$ had a value of $f_{dot}(x_l, x_i)_k$ equal to 0.90 or above.

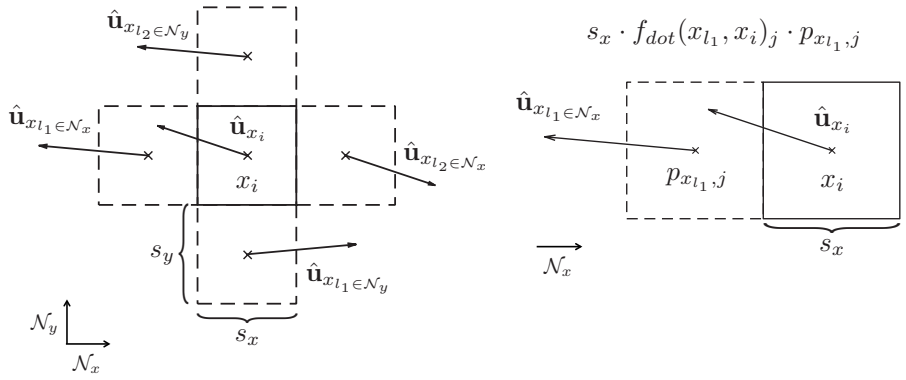
The possibility of anisotropic voxel sizes and slice spacing is considered with the connection-strength factor (s) introduced by Cardoso et al. [28], defined as $s = \{s_x, s_y, s_z\} = \{1/d_x, 1/d_y, 1/d_z\}$ based on real-world distances between the centre of neighbouring voxels. Closer voxels will yield higher weights in the MRF. The total energy (U_{MRF}) (Figure 4.3) for a face-connected neighbourhood $\mathcal{N}_i$ centered at voxel x_i in Eq. 4.12 is defined as

$$\begin{aligned} U_{\mathrm{MRF}}\left(e_k \,|\, p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)}\right) = \sum_{j=1}^{K} \mathbf{G}_{kj} (& \\ & \sum_{l \in \mathcal{N}_i^x} s_x \cdot f_{dot}(l,i)_j \cdot p_{l,j} + \\ & \sum_{l \in \mathcal{N}_i^y} s_y \cdot f_{dot}(l,i)_j \cdot p_{l,j} + \\ & \sum_{l \in \mathcal{N}_i^z} s_z \cdot f_{dot}(l,i)_j \cdot p_{l,j}) \end{aligned} \tag{4.14}$$

where $\mathbf{G}$ represents a K by K matrix whose elements G_{kj} represent the transition energy between tissue classes k and j and the the subscript l denotes the neighbourhood iterator in each direction x, y, z. Since this is a two-class problem, the matrix $\mathbf{G}$ is set up with diagonal elements equal to 0, off-diagonal elements set to 1 and is a constant in our framework.

Initial parameters for the mixture model (Eq. 4.8 and 4.9) are set to $\mu^{(m=0)} = \{0.10, 0.90\}$ on the assumption that both class distributions are significantly different. The class standard deviations are initialised as the original standard deviation of the image filter ($\sigma^{2(m=0)} = \{\sigma^2(\mathcal{F}_t), \sigma^2(\mathcal{F}_t)\}$). Initial values for the MRF energy weights (Eq. 4.12) are set to an even split of 0.5. A termination criteria based on the ratio of likelihood change is set to $\varepsilon = 10^{-3}$.

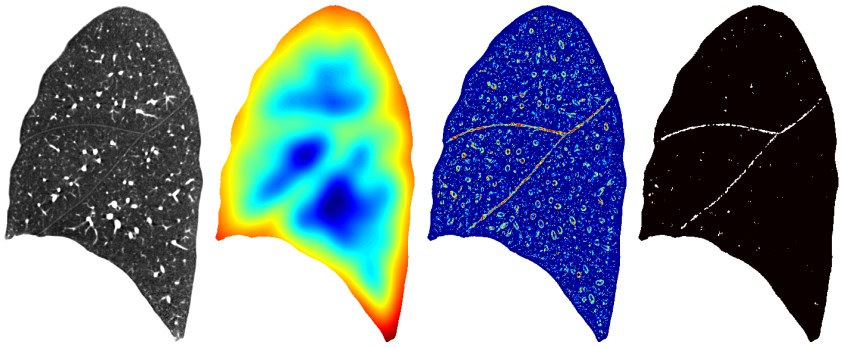
The parameter β in Eq. 4.12 controls the regularisation strength. To mitigate dependence of the segmentation on a user-defined choice, the segmentation is performed iteratively whilst increasing the strength of β. The percentage of high probability fissure voxels ($p_{i,k=1} \geq 0.75$) with respect to the number of voxels n is quantified. This percentage decreases as β rises leading to a fall in false-positive fissure voxels. The initial regularisation



(a) Two-dimensional neighbourhood $\mathcal{N}_i$

(b) Features used in the MRF energy for a component (l_1) in $\mathcal{N}_i^x$

Figure 4.3: Illustration of the local neighbourhood $\mathcal{N}_i$ used in Eq. 4.14. The index l iterates over each component in $\mathcal{N}_i$ whilst j is the tissue class iterator. The real-world distances are represented by s_x and s_y, both measured in mm. © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.



(a) $\mathcal{I}_t$ **(b)** π_t **(c)** $\mathcal{F}_t$ **(d)** $\mathcal{P}_{t,\text{fissure}}$

Figure 4.4: Given a patient $\mathcal{I}_t$ (a), an anatomical prior (π_t) (b) is derived from the vessel density and airway tree distance transform. A multi-scale filter is applied to $\mathcal{I}_t$ to yield $\mathcal{F}_t$ (c). Gaussian mixture modelling with an MRF yields a probabilistic segmentation (d) of the fissures $\mathcal{P}_{t,\text{fissure}}$. © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.

is $\beta = 0.75$. The EM framework is run and the percentage high probability voxels is quantified. If there is a convergence of this percentage, the segmentation framework terminates. Otherwise, the regularisation is automatically increased by 0.50. Convergence is defined when the percentage has not fallen by at least 2% in 5 successive iterations. The output of the framework after convergence yields probabilistic fissure ($\mathcal{P}_{t,\text{fissure}}$) and tissue ($\mathcal{P}_{t,\text{tissue}}$) maps for each image $\mathcal{I}_t$. To obtain a binary segmentation of the fissures ($\mathcal{S}_t$), a two-pass analysis is performed. A connected component analysis of $\mathcal{P}_{t,\text{fissure}}$ using a face-connected neighbourhood is performed. Firstly, all components with a median probability below 0.50 are removed then all components below a volumetric threshold of 0.50mL are discarded.

4.2.3 Groupwise fissure prior

In a given patient cohort, there will be a range of cases with incomplete fissures, which will complicate the segmentation of the lobes. The goal is to combine all segmented fissures into a groupwise space to create an average fissure model to help guide the lobe segmentation in problematic cases.

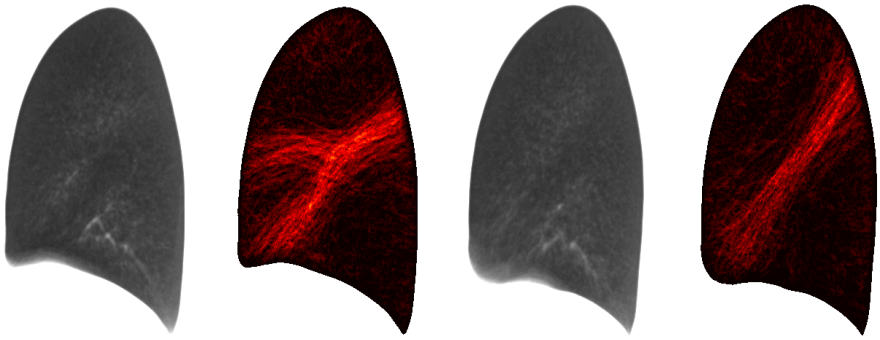
4.2.3.1 Groupwise registration

Given a set of T patients $\mathcal{I}=\{\mathcal{I}_1,\mathcal{I}_2,\ldots,\mathcal{I}_T\}$, a common average space $\Omega_{\hat{\mathcal{I}}}$ is computed by iteratively registering the set of patients $\mathcal{I}$ to the Frêchet mean. The output is an average image $\hat{\mathcal{I}}_{avg}$ and a set of forward and backward transformations ($\varphi_{\mathcal{I}_T,\hat{\mathcal{I}}}$ and $\varphi_{\hat{\mathcal{I}},\mathcal{I}_t}$) such that $\varphi_{\mathcal{I}_t,\hat{\mathcal{I}}} : \Omega_{\mathcal{I}_t} \rightarrow \Omega_{\hat{\mathcal{I}}}$ and $\varphi_{\hat{\mathcal{I}},\mathcal{I}_t} : \Omega_{\hat{\mathcal{I}}} \rightarrow \Omega_{\mathcal{I}_t}$.

The registration is performed using the NiftyReg software package [139]. The algorithm is motivated by the work of Ashburner et al. [8]. All patients in $\mathcal{I}$ are initially registered to an initial template image $\hat{\mathcal{I}}_{avg}^{(n=0)}$, which is chosen at random from the set of images $\mathcal{I}$. The initial average space is created using a rigid registration. This prevents the atlas from being biased by the geometry of the initial template image. A set of n_1 affine registrations using symmetric block-matching [135] are then performed followed by a set of n_2 non-rigid registrations. The non-rigid registration uses a stationary velocity field, which is parameterised by a cubic B-spline with a 12mm spacing. The locally normalised cross-correlation is used as the similarity with a Gaussian kernel of 50mm. The number of iterations was determined in a pilot experiment by computing the sum of squared differences similarity between successive average images $\hat{\mathcal{I}}_{avg}^{(n)}$ and $\hat{\mathcal{I}}_{avg}^{(n+1)}$ at iterations n and $n+1$. Convergence of similarity measures occurred after $n_1 = 5$ and $n_2 = 5$ affine and nonrigid registrations.

At each iteration (n) of the algorithm, all patients are registered to the average im-

age $\hat{\mathcal{I}}_{avg}^{(n-1)}$. In order to create the new average image $\hat{\mathcal{I}}_{avg}^{(n)}$ and the new space $\Omega_{\hat{\mathcal{I}}^n}$, the average transformation from all patients is computed in the log-Euclidean space, which is used to demean each patient transformation. All patients are subsequently resampled using the demeaned transformations to create $\hat{\mathcal{I}}_{avg}^{(n)}$. This new average image $\hat{\mathcal{I}}_{avg}^{(n)}$ is obtained by averaging all the resampled images.

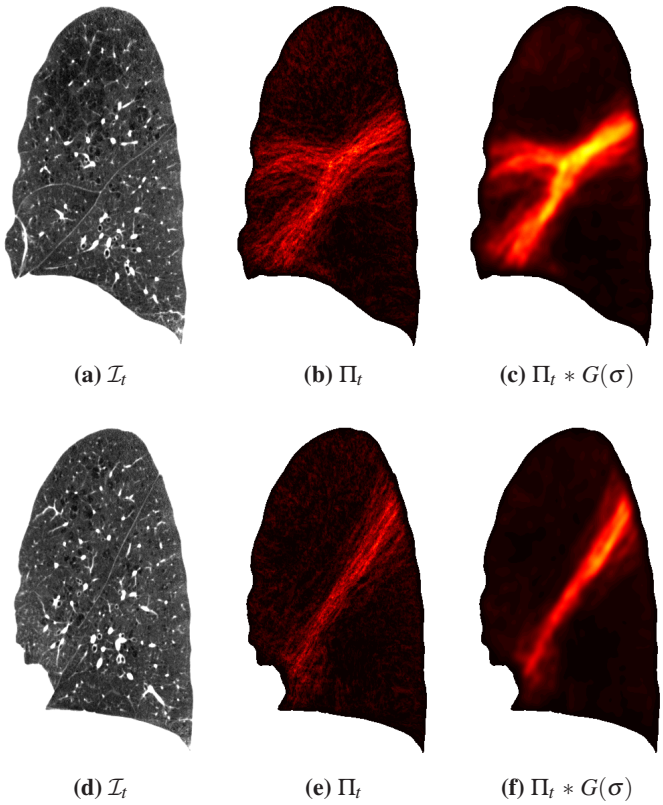


(a) $\hat{\mathcal{I}}$ **(b)** $\hat{\mathcal{S}}_{\text{avg}}$ **(c)** $\hat{\mathcal{I}}$ **(d)** $\hat{\mathcal{S}}_{\text{avg}}$

Figure 4.5: Average fissure in the groupwise space $\Omega_{\hat{\mathcal{I}}}$. The average of the right (a-b) and left (c-d) lung with the respective average fissures can be seen. © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.

4.2.3.2 Construction of the population prior

The set of patients $\mathcal{I}=\{\mathcal{I}_1,\mathcal{I}_2,\ldots,\mathcal{I}_T\}$ will have a corresponding set of binary fissure volumes $\mathcal{S}=\{\mathcal{S}_1,\mathcal{S}_2,\ldots,\mathcal{S}_T\}$. The transformations resulting from the groupwise registration ($\varphi_{\mathcal{I}_t,\hat{\mathcal{I}}}$ and $\varphi_{\hat{\mathcal{I}},\mathcal{I}_t}$) are exploited to build the fissure prior ($\hat{\mathcal{S}}_{\text{avg}}$). Each fissure segmentation $\mathcal{S}_t$ is resampled into the groupwise space ($\Omega_{\hat{\mathcal{I}}}$) using the respective forward transformation $\varphi_{\mathcal{I}_t,\hat{\mathcal{I}}}$ to yield $\hat{\mathcal{S}}_t$. All resampled fissures ($\hat{\mathcal{S}}=\{\hat{\mathcal{S}}_1,\hat{\mathcal{S}}_2,\ldots,\hat{\mathcal{S}}_T\}$) are averaged in the groupwise space to create the average fissure $\hat{\mathcal{S}}_{\text{avg}}$ (Figure 4.5). In order to exploit this information to help segment the lobes, the average fissure is resampled using the backwards transformation $\varphi_{\hat{\mathcal{I}},\mathcal{I}_t}$ into each patient space ($\Omega_{\mathcal{I}_t}$) resulting in Π_t. This prior is normalised to the range [0,1] for each patient and is subsequently smoothed using a Gaussian kernel ($\sigma = 2.5$mm). This produces a prior in the space of each patient, denoting a region where the fissure is expected (Figure 4.6).



(a) $\mathcal{I}_t$ **(b)** Π_t **(c)** $\Pi_t * G(\sigma)$

(d) $\mathcal{I}_t$ **(e)** Π_t **(f)** $\Pi_t * G(\sigma)$

Figure 4.6: Groupwise prior in the patient space $\Omega_{\mathcal{I}_t}$. a) The patient volume ($\mathcal{I}_t$), b) the resampled average fissure (Π_t) and c) the smoothed prior ($\Pi_t * G(\sigma)$). Within this example, the right lung can be viewed in (a), (b) and (c) whilst the left lung is in (d), (e) and (f).

4.2.4 Watershed surface fitting

The segmented fissure, the groupwise fissure prior and anatomical information are combined into a cost image inspired by the formulation of Lassen et al. [109]. They created a cost function by combining information from the vessel and airway tree, the segmented fissure and the voxel intensities. I built on this work by extending the cost function to utilise a fissure groupwise prior derived from the population that needs to be analysed.

4.2.4.1 Watershed cost function

The population prior (Π_t) is first combined with the segmented fissure ($\mathcal{S}_t$). The aim of this step is to produce an initial cost function, with regions of complete and incomplete fissures accentuated using information from the segmentation $\mathcal{S}_t$ and the population prior Π_t. The inverted Euclidean distance function is applied to $\mathcal{S}_t$ to help deal with minor gaps in the segmentation. It is normalised with min-max scaling to the range [0,1] with a value of 1 at the fissure. Only regions in the distance map ($f_{dist}(\mathcal{S}_t)$) within 2.5mm of the fissure are considered. The distance map and the population prior are averaged and convolved with a small Gaussian kernel ($\sigma = 1.0$mm) to produce a smooth map in Eq. 4.15. The magnitude of c_1 will be strongest when $f_{dist}(\mathcal{S}_t)$ and Π_t are in the same anatomical location. When there are large gaps in $\mathcal{S}_t$ due to fissure incompleteness, Π_t will provide a local maxima.

$$c_{t,1} = \left(\frac{f_{dist}(\mathcal{S}_t) + \Pi_t}{2} \right) * G(\sigma)\,. \tag{4.15}$$

This is then combined with the anatomical information ($\pi_{t,ik=1} \forall i \rightarrow \pi_{t,k=1}$) and the binary segmentation ($\mathcal{S}_t$):

$$\mathcal{C}_t = \left(\frac{c_{t,1} + \pi_{t,k=1} + \mathcal{S}_t}{3} \right) \tag{4.16}$$

The fissure likelihood based on the vessel and airway tree ($\pi_{t,k=1}$) provides a satisfactory estimate for regions of low and high fissure probability and helps guide the segmentation into regions of low vessel and airway density. In addition to c_1, the original segmentation ($\mathcal{S}_t$) is reintroduced. This is performed to produce a high value in $\mathcal{C}_t$ at the segmented fissure since this is the true location of the lobar border.

4.2.4.2 Lobe seed labelling

I used the method used by Doel et al. [54] to generate initial lobe seed labels from the segmented airway tree. The seed labels are dilated and allowed to grow according to the

vessel density map for a limited amount of iterations using the watershed algorithm.

4.2.4.3 Lobe seed labelling post-processing

The segmentation is dependent on the initial seed labels. The quality of the initial labelling can be affected by segmentation failures and errors in the bronchial labelling. If a lobe seed is incorrectly labelled, the resulting segmentation will be erroneous despite successful segmentation of the fissures. To improve the robustness of the pipeline with respect to the airway tree segmentation, the following errors are accounted for: 1) mislabelled branches and 2) unsegmented branches.

A. Seed label correction - labelling errors

To detect and remove mislabelled branches (Figure 4.7), the centre of mass (CoM) of each set of label seeds is quantified. For each label set, the intra-label distance of each component to the label CoM is computed. The inter-label distance of the components to all other label CoMs are also quantified. The dilated components are removed iteratively to minimise the amount of deleted seeds. A seed is marked for removal if its inter-label distance is smaller than its intra-label distance. The seed with the smallest inter-component distance of all candidate components is marked for removal. This component is discarded if its removal does not cause the number of components for that label to fall below a threshold ($r_{threshold} = 4$). If this threshold is met, the component with the next smallest distance is considered. Once a component is removed, the above method is repeated until removal is no longer possible. This enforces maximum separability between the seeds and removes all potentially erroneously labelled seeds. An example of the algorithm can be viewed in Figure 4.8.

B. Seed label correction - airway segmentation errors

If the airway tree segmentation fails (Figure 4.9), labelling of the branches will not yield the necessary seeds to segment all lobes. In this instance, the anatomical information ($\pi_{t,k=1}$) and its distribution in non-fissure regions (4.17) is considered. The threshold is considered by analysing the distribution of the prior ($\pi_{t,k=1}$) at the fissure $\mathcal{S}_t$:

$$\pi_{t,threshold} = \mu\left(\pi_{t,k=1} \in \mathcal{S}_t\right) - 2\sigma\left(\pi_{t,ik=1} \in \mathcal{S}_t\right) \tag{4.17}$$

The low minima regions of $\pi_{t,k=1}$ are defined as $\pi_{t,k=1} < \pi_{t,threshold}$. These correspond to

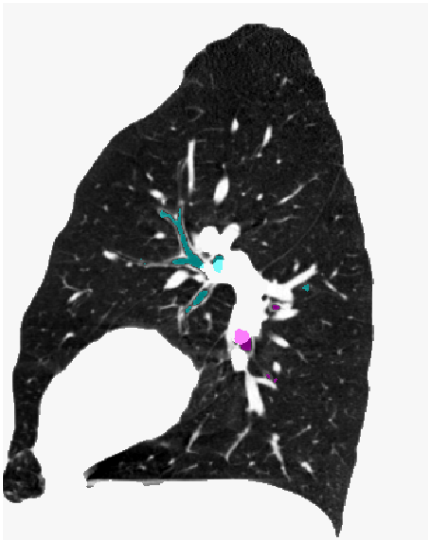
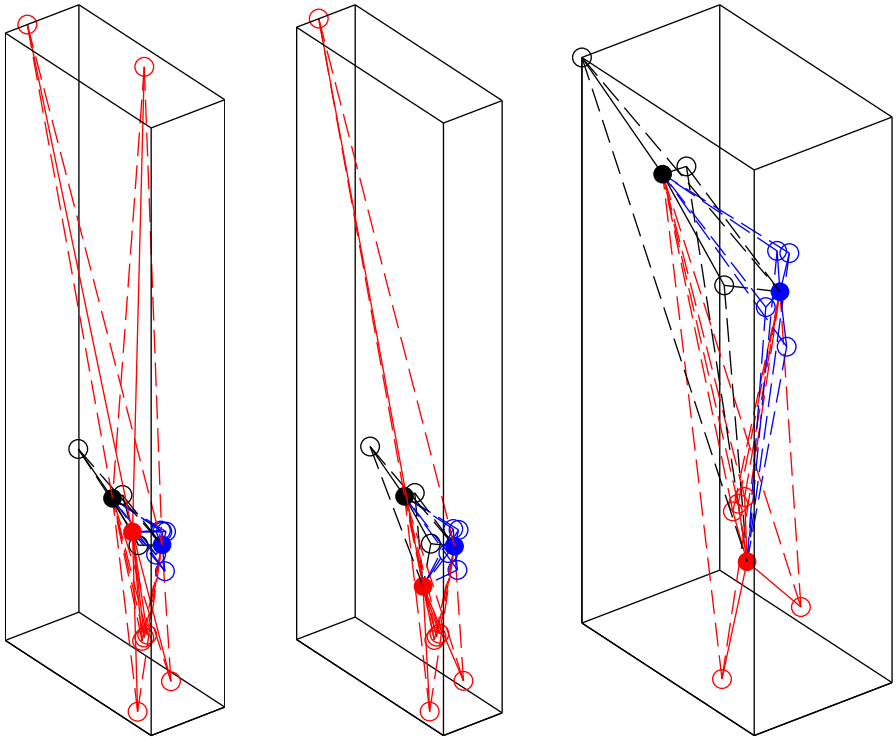


Figure 4.7: Labelling error in the lobe seeds in the left lung. A lower left lobe seed has been incorrectly labelled as upper left lobe.



(a) Iteration 1 **(b)** Iteration 2 **(c)** Iteration 3

Figure 4.8: Schematic of the seed label correction algorithm. There are labelled components for the upper right lobe (black), middle right lobe (blue) and lower right lobe (red). The centre of mass of the components per lobe are denoted by the filled circles whilst the centre of mass of each component is an unfilled circle. The solid lines denote the connections that calculate intra-label distances whilst the dashed lines represent the inter-label distances. As the algorithm iterates, two erroneous lower right lobe seed label components (top of figure) are automatically removed in (b) and (c) causing the centre of mass of the lower lobe component to move away from the other lobes.

regions of high vessel and airway branching density. A mask of these regions is created and a connected component analysis is performed to extract the regions of local minima (Figure 4.10). These regions are analysed based on their position within the lung and are exploited to generate new seed labels should the airway tree segmentation fail.

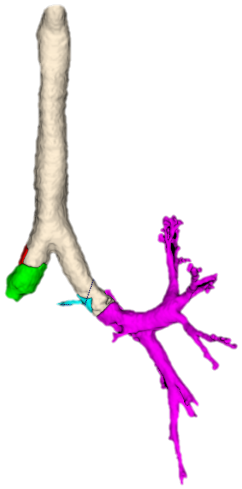
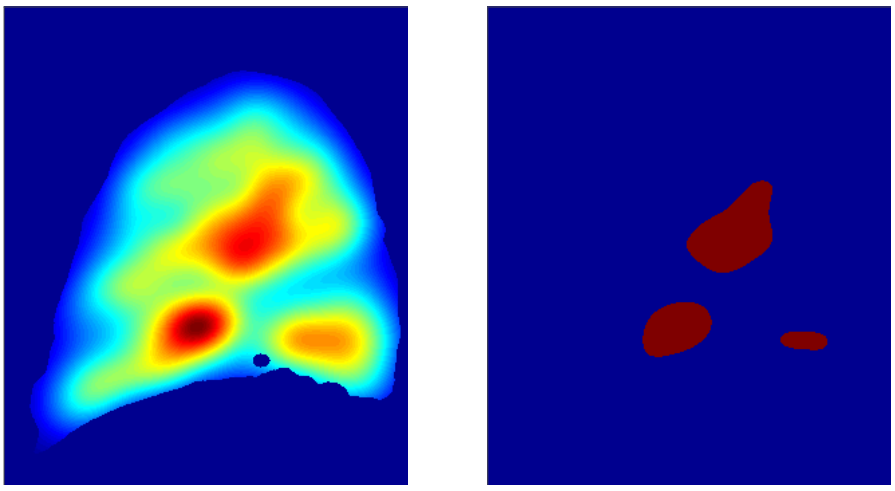


Figure 4.9: Three-dimensional reconstruction of a segmented airway missing the branches in the right lung.

C. Final lobe segmentation

Segmentation of the lobes ($\mathcal{L}_t$) is obtained by combining the cost image $\mathcal{C}_t$ and the processed seed labels in a watershed segmentation. The lobar boundaries are smoothed to deal with minor artefacts in the segmentation. This is performed by normalised convolution with a 4.0mm^3 Gaussian kernel.



(a) Vessel density map **(b)** Binary mask after thresholding

Figure 4.10: Seed label generation by analysis of the vessel density map for the airway tree in Figure 4.9. a) Analysis of the vessel density map using the fissure segmentation with Equation 4.17 facilitates a data-driven way to threshold the vessel density map into regions of low vessel density. b) The creation of a binary mask using the thresholding allows one to generate candidate regions in the lung for the lobe seeds in cases with no segmented airway branches.

4.3 Data

4.3.1 Dataset 1

Dataset 1 was based on the LObe and Lung Analysis 2011 (LOLA11) challenge [99]. It consists of 55 volumetric chest CT scans originating from a variety of source with a range of scans containing serious pathology and abnormalities. The inplane resolution is between 0.53mm and 0.78mm whilst the slice thickness ranges from 0.3 to 1.5mm. The organisers manually segmented the lobes on 9 coronal slices with two human observers and were instructed only to label when the boundaries were visible. The inter-observer agreement between the lobar borders was 1.50mm $\pm$1.28mm.

4.3.2 Dataset 2

Dataset 2 was used to quantitatively and qualitatively validate the framework on patients with COPD from the COPDGene study [167]. I created a quantitative ($N_{quant} = 30$) and qualitative set ($N_{qual} = 50$) by randomly selecting patients from the study. Minimum thresholds ($> 10\%$) for the level of emphysema in the inspiration scan and gas trapping in the corresponding expiration scan were set in the qualitative cohort to capture cases with sig-

nificant pathology. The quantitative set averaged $12.8\% \pm 11.12\%$ emphysema whilst the qualitative set averaged $22.50\% \pm 3.60\%$ emphysema.

Analysed scans stem from GE Medical Systems (LightSpeed 16, Lightspeed VCT), Siemens (Sensation 16, Sensation 64 and Definition) and Philips (Brilliance 64) scanners. Scans with the STANDARD (GE), AS+ B31f and B31f (Siemens), and 64 B (Philips) reconstruction algorithms were analysed. Information about the scanning protocols can be viewed at the COPDGene website[2]. The slice thickness of the scans range from 0.62mm to 1.00mm with in-plane dimensions ranging from 0.52 to 0.90mm.

The quantitative cohort was built by manually tracing the fissures in every fifth sagittal slice using ITK-SNAP [252]. The radiologist was asked to manually trace the fissures using three labels. Label 1 was used when the fissures were visible. Label 2 was employed in cases where extrapolation was possible. Label 3 was used in areas of high fissure uncertainty.

The manual segmentation provided an approximate estimate of fissure incompleteness with an average of $12.4\% \pm 8.3\%$ across the quantitative set. This was computed by considering the percentage of voxels labelled 2 and 3. Intra-observer variability was obtained by a repeated segmentation of 3 datasets with varying degrees of fissure incompleteness (6.7%, 23.0% and 31.3%). These were performed 14 days after to minimise recall bias. The intra-observer agreement for all lobar boundaries across all patients was $1.54\text{mm} \pm 0.45\text{mm}$.

4.4 Experiments and results

4.4.1 Fissure segmentation evaluation

I investigated the effects of parameters A and B (Eq. 4.3) and the performance of our segmentation framework on the quantitative set of dataset 2 using label 1 of the reference set.

I used the method presented by Xiao et al. [242] to evaluate our fissure segmentation. I did not define a volume of interest (VoI) using a 40mm width band around each reference as this ignores potential false positives in the validation. The F_1-score was used as quantitative index of performance. It is defined as $2 \cdot (Precision \cdot Recall)/(Precision + Recall)$. The magnitude of F_1 reflects the similarity between the segmentation and the reference. *Precision* and *Recall* are defined respectively as $TP_1/(TP_1 + FP)$ and $TP_2/(TP_2 + FN)$ [242]. *Precision* was quantified by considering the overlap of the binary result ($\mathcal{S}$) with

[2] copdgene.org/sites/default/files/COPDGene%20MOP%2006.19.2009.pdf

the reference. A 3mm tolerance band was defined around the reference segmentation as performed by Xiao et al. [242]. Voxels of $\mathcal{S}$ are classified as true positive (TP_1) if they fall within the 3mm band and false positive (FP) if otherwise. *Recall* was computed by defining a 3mm band around the binary result ($\mathcal{S}$). Reference voxels within this band were classified as TP_2 and those outside as false-negative (FN).

I segmented the fissures using parameters $A, B \in [0.05, 0.50]$. I illustrate the performance of the segmentation for a subset of the parameter $B \in [0.05, 0.10, 0.15, 0.25]$ with $A \in [0.05, 0.50]$ in Figure 4.11. The best performance over all datasets was achieved with parameters $A = 0.25$ and $B = 0.10$ with a median F_1-score of 0.90 with median False-Discovery Rate and False-Negative-Rate of 0.08 and 0.13 respectively. The F_1-score remained relatively stable when set in the range $A^* = B^* \in [0.10, 0.35]$ (Figure 4.11). The mean F_1-score over all combinations ($A^* \times B^*$) was 0.87 ± 0.02 demonstrating stability in algorithm performance. The mean F_1 over all values of A for increasing values of B is 0.87 ± 0.03 (Figure 4.11a), 0.87 ± 0.03 (Figure 4.11b), 0.86 ± 0.04 (Figure 4.11c), 0.83 ± 0.08 (Figure 4.11d). At higher values of B, this drops to 0.82 ± 0.09 ($B = 0.30$), 0.81 ± 0.10 ($B = 0.35$), 0.79 ± 0.12 ($B = 0.40$), 0.78 ± 0.14 ($B = 0.45$) and 0.77 ± 0.15 ($B = 0.50$). This is expected as higher values decreases the separation between the tissue and fissure-class distributions.

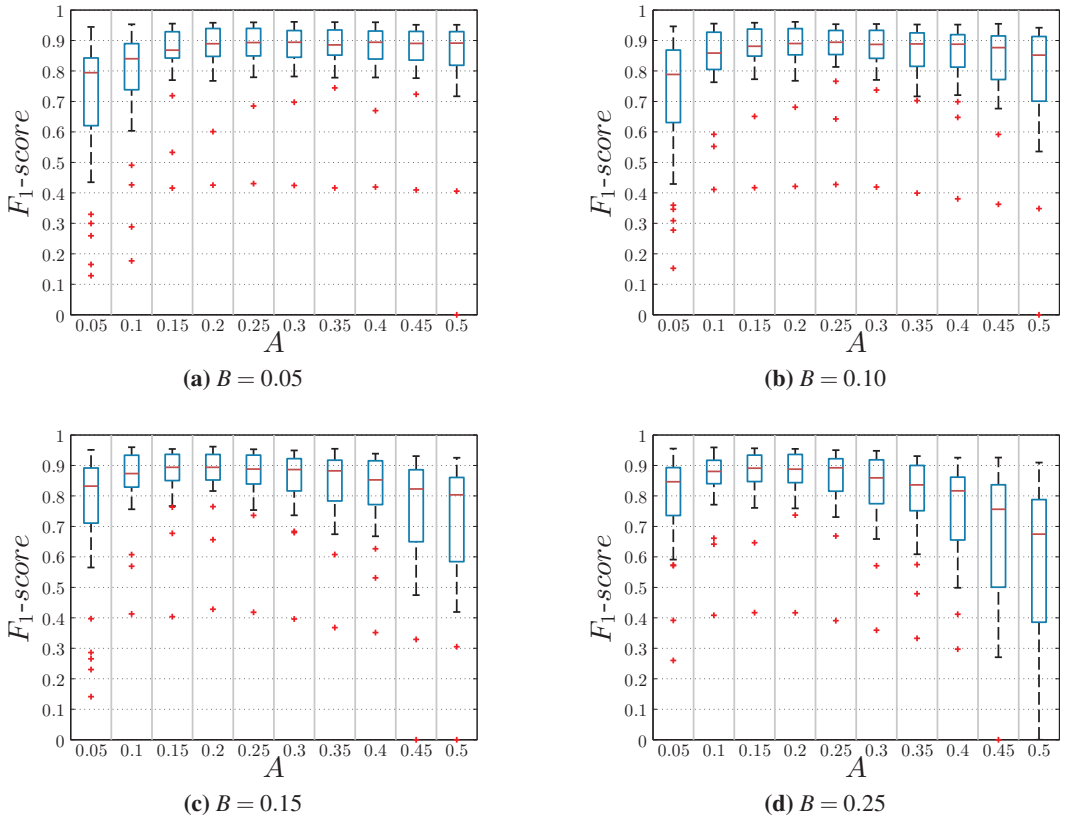
4.4.2 Lobe segmentation validation

Algorithm parameters quoted within Section 4.2 were used in the validation of dataset 1 and 2. The fissure filter parameters used were $A = 0.20$ and $B = 0.20$.

4.4.2.1 Dataset 1 - LOLA11

I evaluated our algorithm on the LOLA11 cohort and submitted our results for evaluation [99]. The LOLA11 evaluation metric is the volume overlap between the submission and the reference segmentation of one observer. The organisers defined a 2mm slack border around the borders of the lung and lobes to account for inter-observer variability. Voxels within this border were not accounted for during evaluation. The overlap is calculated for each lobe across all patients. I report the mean $\pm$ standard deviation, first quartile (Q1), median and third quartile (Q3) of the scores across all 55 patients. The LOLA score is calculated as the average of all average overlaps over all lobes. Table 4.1 shows the score for our lobe segmentation and those of van Rikxoort et al. [220] and Lassen et al. [109].

Five algorithms have been validated for lobe segmentation on this cohort. However, I



(a) $B = 0.05$ **(b)** $B = 0.10$

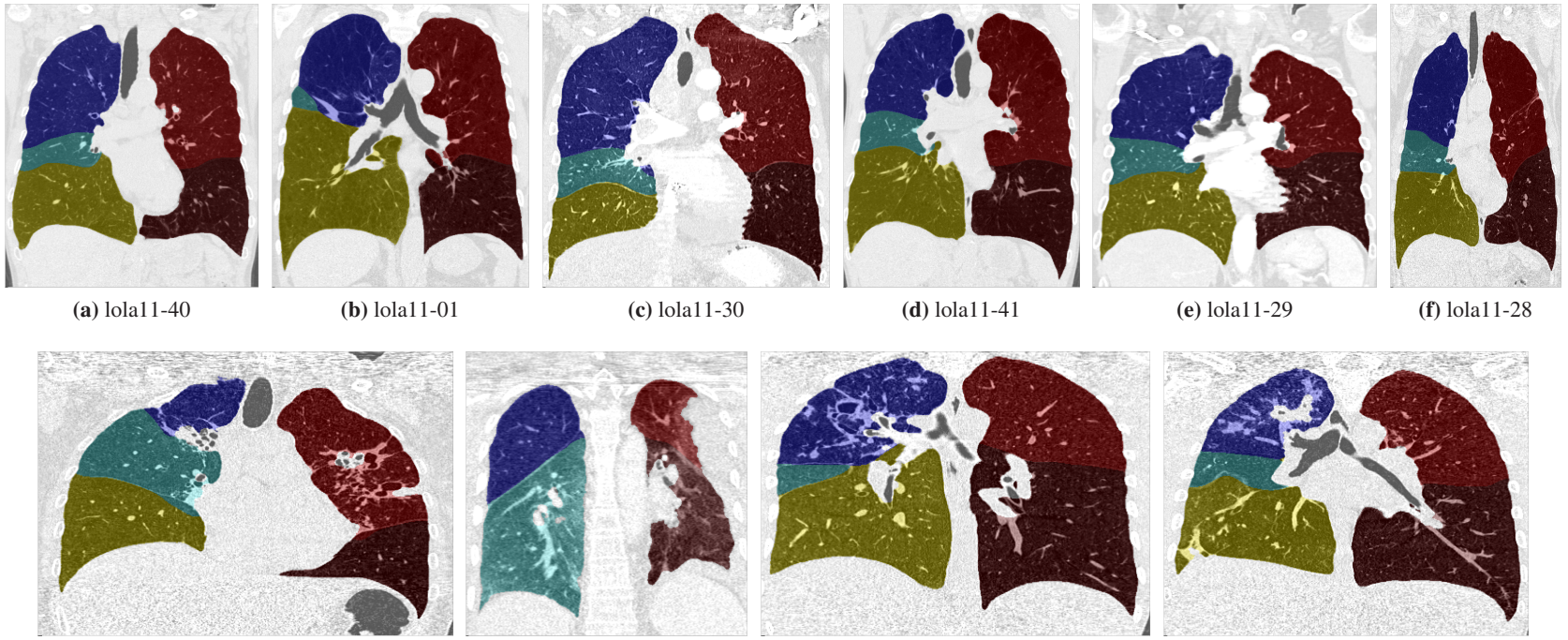
(c) $B = 0.15$ **(d)** $B = 0.25$

Figure 4.11: Boxplots of the fissure segmentation with increasing values of B whilst varying constant A of the multiscale filter (Eq. 4.3). © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.

have restricted our comparison to van Rikxoort et al. [220] and Lassen et al. [109] as they are fully automatic and do not require interactive post-processing to correct segmentations. I achieved the highest automatic average lobe score of 0.884 and the second highest average median overlap of 0.950 (Table 4.1).

Table 4.1: LOLA 11 lobe segmentation results. LUL=left upper lobe, LLL=left lower lobe, RUL=right upper lobe, RML=right middle lobe and RLL=right lower lobe.

Lobe	Mean ± STD	Q1	Median	Q3
LUL	0.906 ± 0.202	0.946	0.975	0.988
LLL	0.880 ± 0.243	0.919	0.962	0.980
RUL	0.928 ± 0.071	0.888	0.960	0.980
RML	0.799 ± 0.235	0.759	0.891	0.941
RLL	0.908 ± 0.194	0.937	0.961	0.976
Our method	0.884		0.950	
Lassen et al. [109]	0.881		0.951	
van Rikxkoort et al. [220]	0.851		0.943	



(a) lola11-40 **(b)** lola11-01 **(c)** lola11-30 **(d)** lola11-41 **(e)** lola11-29 **(f)** lola11-28

(g) lola11-55 **(h)** lola11-21 **(i)** lola11-14 **(j)** lola11-25

Figure 4.12: Illustration of the segmentation on a variety of cases from dataset 1 with complete and incomplete fissures in addition to various levels of pathology. © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.

4.4.2.2 Dataset 2 - Quantitative COPDGene

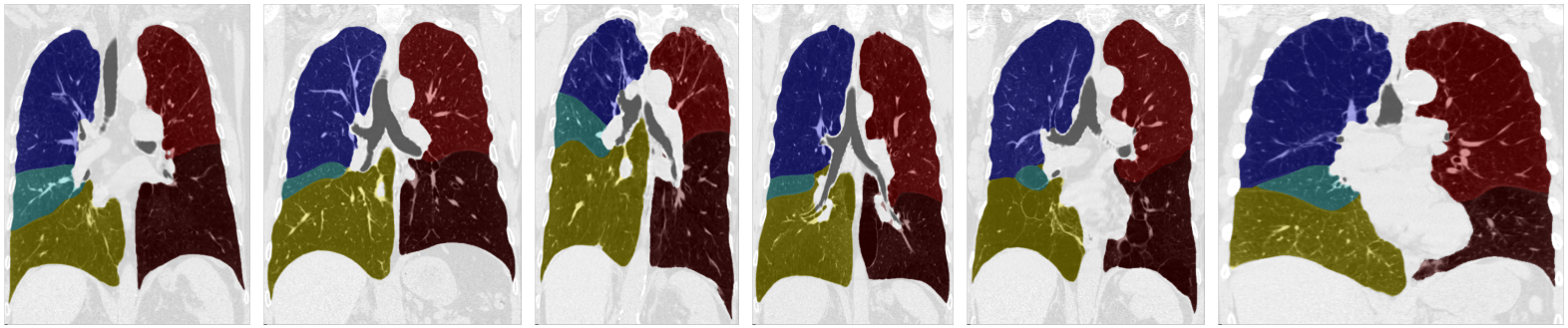
I assessed the performance of our algorithm quantitatively on dataset 2 using the mean, maximum and root-mean square error (RMSE) distance from the manual reference to the automatic boundary. This was performed by calculating the three-dimensional Euclidean distance between the reference voxels and the closest point on the automatic segmentation.

I assessed the segmentation for each label (Table 4.2). The algorithm achieved a mean of 2.01mm $\pm$ 6.24mm when the fissures were visible (label 1). In cases where fissure extrapolation was possible (label 2), a mean of 5.16mm $\pm$ 6.12mm was achieved. The performance dropped to 7.31mm $\pm$ 4.88mm in regions of highest uncertainty (label 3). In one case, the right lobe segmentation failed and in a second case, segmentation of the right and left lobes failed. The failure was due to major errors in the airway branching labelling, which could not be corrected using our methodology. With a 2mm slack border, the mean distances were 1.65 ± 3.28mm (label 1), 3.31 ± 5.93mm (label 2) and 6.18 ± 4.70mm (label 3). Errors were due to instances where the groupwise fissure was significantly different from label 2 and 3 of the reference or slightly biased the segmentation of label 1. Further typical errors were in cases where emphysametous bullae appearing as fissures lead to isolated errors in the lobe boundary segmentation.

The performance of my algorithm was also assessed against approximate fissure incompleteness. The correlation between the mean error (distance to the reference) and fissure incompleteness was calculated. The relationship between the standard deviation of the errors and fissure incompleteness was also calculated. This was performed with Pearson's correlation coefficient (ρ) and processed for each lobar boundary. Weak and moderate relationships were observed between the mean distance and the degree of fissure incompleteness for each boundary (right minor: $\rho = 0.47\,(p < 0.05)$, right major: $\rho = 0.66\,(p < 0.05)$ and left major: $\rho = 0.35\,(p > 0.05)$). Similar findings were observed in the standard deviation (right minor: $\rho = .46\,(p < 0.05)$, right major: $\rho = 0.61\,(p < 0.05)$ and left major: $\rho = 0.23\,(p > 0.05)$).

Table 4.2: Quantitative Results - Fissure Metrics. © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.

		average of patient mean (mm)	average of patient max (mm)	average of patient rmse (mm)
Label 1	Right Major	1.66 ± 3.61	15.42 ± 14.91	2.72 ± 5.67
	Right Minor	2.31 ± 11.10	18.60 ± 20.10	3.32 ± 12.82
	Left Major	2.07 ± 4.03	28.50 ± 27.89	3.95 ± 6.85
	Total	2.01 ± 6.24	20.85 ± 20.96	3.33 ± 8.44
Label 2	Right Major	5.44 ± 5.55	14.22 ± 14.45	6.69 ± 6.98
	Right Minor	5.39 ± 8.50	18.65 ± 18.96	7.23 ± 8.48
	Left Major	4.65 ± 4.30	16.15 ± 16.75	6.01 ± 5.81
	Total	5.16 ± 6.12	16.34 ± 16.72	6.64 ± 7.09
Label 3	Right Major	7.21 ± 4.18	20.95 ± 21.02	8.65 ± 6.15
	Right Minor	7.56 ± 5.49	19.05 ± 17.62	8.44 ± 7.28
	Left Major	7.16 ± 34.96	23.23 ± 23.23	9.78 ± 5.58
	Total	7.31 ± 4.88	21.08 ± 20.62	8.95 ± 6.34



(a) Qual. 1 **(b)** Qual. 2 **(c)** Qual. 3 **(d)** Qual. 4 **(e)** Qual. 5 **(f)** Qual. 6

(g) Quant. 1 **(h)** Quant. 2 **(i)** Quant. 3 **(j)** Quant. 4 **(k)** Quant. 5 **(l)** Quant. 6

Figure 4.13: Illustration of the segmentation on examples from the qualitative set of dataset 2 and the quantitative set (bottom row) at various levels of COPD severity.

4.4.2.3 Dataset 2 - Qualitative COPDGene

The radiologist qualitatively assessed (Table 4.3) my algorithm on 50 patients with advanced disease using an adapted scoring system of van Rikxoort et al. [222], which scores the segmentations out of five. The radiologist assessed each lobe segmentation on the sagittal plane. The highest score (5) corresponded to a segmentation error below 3mm. A score of 4 reflected a segmentation error at any location between 3mm and 12mm. A score of 3 reflected a segmentation error greater than 12mm but where the overall lobe segmentation remained acceptable for analysis. The lowest scores (2 and 1) were awarded when the maximum segmentation error was greater than 12mm and segmentation quality was either equivocal or unusable. Scores were assigned to both complete and incomplete fissures.

The algorithm showed good performance across complete (3.9 ± 0.3) and incomplete fissures (3.8 ± 0.5). There was one notable failure (score = 2) to segment the right major fissure in a case with a complete fissure and one failure (score = 1) of a segmentation of a right major fissure when it was grossly incomplete. Across cases with complete and incomplete fissures, the most commonly awarded score was 4. Only a small proportion of cases had maximum errors $<$ 3mm (22% and 12.5% for complete and incomplete right minor fissures whilst 0% and 7.3% for complete and incomplete right major fissures).

4.4.3 Effect of the groupwise prior

I assessed the performance of the algorithm with and without the groupwise prior (Π) on dataset 2. I computed the mean of the distances from the automated segmentation to the reference and compared this to results using the prior. I did not include the boundaries that failed in the quantitative analysis of Section 4.4.2.2 as these boundaries also failed without using the groupwise prior. I omitted cases with minor fissure incompleteness in the analysis. This was defined when a lobar boundary had less than 1% fissure incompleteness. This led to 19 analysed patients for the right major fissure, 23 for the right minor and 19 for the left major. There was $9.72\% \pm 8.66\%$ fissure incompleteness in the right major, $36.44\% \pm 18.91\%$ for the right minor and $11.61\% \pm 9.81\%$ for the left major in the new cohort. We calculated the cohort average for each boundary and for the segmentation labels of fissures not visible in CT, delineated as label 2 and 3. I performed a two-sample t-test under the null hypothesis that the mean results of the segmentation with and without the groupwise prior are significantly different.

I found a general increase in the distance to the reference for label 2 ($5.87\text{mm} \pm$

Table 4.3: Qualitative COPDGene results. © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.

(a) Complete fissures (label 1)

Score	Right Major	Right Minor	Left Major	Total Score
% 5	0.0	22.0	0.0	
% 4	82.4	58.0	76.5	
% 3	15.6	20.0	23.5	3.9 ± 0.3
% 2	2.0	0.0	0.0	
% 1	0.0	0.0	0.0	

(b) Incomplete fissures (label 2 and 3)

Score	Right Major	Right Minor	Left Major	Total Score
% 5	7.3	12.5	2.4	
% 4	68.3	50.0	70.7	
% 3	22.0	37.5	26.8	3.8 ± 0.5
% 2	0.0	0.0	0.0	
% 1	2.4	0.0	0.0	

3.72mm to 7.60mm $\pm$ 6.49mm) and label 3 (7.10mm $\pm$ 3.67mm to 8.59mm $\pm$ 5.77mm) when excluding the groupwise prior. I did not find a significant difference between the sets of mean distances for each lobar boundary stratified by reference label (Table 4.4). This is due to the fact there may be extreme differences due to failures without the prior, smaller improvements using the prior but also cases where the prior negatively affects extrapolation of the fissure.

In areas of significant fissure incompleteness, the groupwise prior may help avoid leaking of the seed labels during the surface fitting whilst guiding the segmentation to the most probable location based on the population and the patient anatomy. This occurred in several cases (Figure 4.14 a and b) where either the left major border or the right major border failed without the prior. Within this cohort, the right minor fissure had the highest level of fissure incompleteness. In various cases (Figure 4.14 c and d), the prior helped drive the lobar border towards the reference. However, there are several modes of variation in the right minor fissure (Figure 4.5). The patient anatomy may differ greatly from the population mean. The prior may negatively affect the final segmentation in areas of incomplete

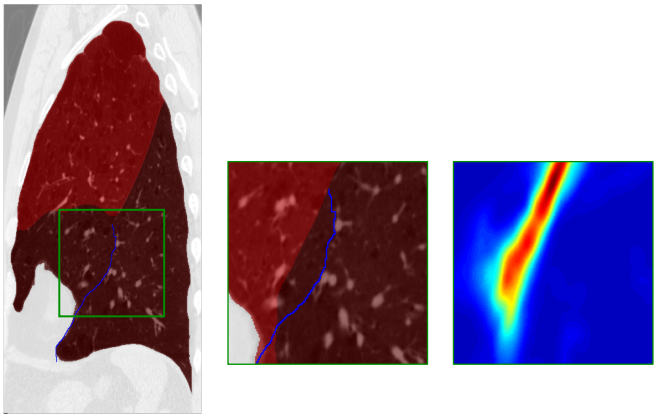
fissures. This led to a smaller difference in the population means in the right minor fissure (6.28mm to 6.51mm in label 2 and 6.41mm to 6.50mm in label 3). Despite this limitation, I can conclude that the groupwise fissure prior, constructed from the same cohort using a combination of complete and incomplete segmented fissures is advantageous.

Table 4.4: Quantitative Results - Fissure Metrics with and without Groupwise Prior. © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.

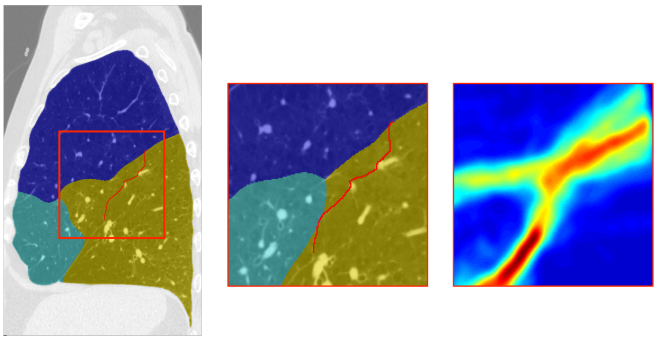
		WITH (mm)	WITHOUT (mm)	p-Value
Label 2	Right Major	5.29 ± 3.00	5.71 ± 2.75	p=0.65
	Right Minor	6.28 ± 3.52	6.51 ± 3.51	p=0.83
	Left Major	6.05 ± 4.64	10.59 ± 16.32	p=0.25
	Total	5.87 ± 3.72	7.60 ± 6.49	-
Label 3	Right Major	6.94 ± 3.24	8.18 ± 4.28	p=0.34
	Right Minor	6.41 ± 2.81	6.50 ± 2.84	p=0.92
	Left Major	7.94 ± 4.95	11.09 ± 10.19	p=0.67
	Total	7.10 ± 3.67	8.59 ± 5.77	-

4.5 Discussion and conclusion

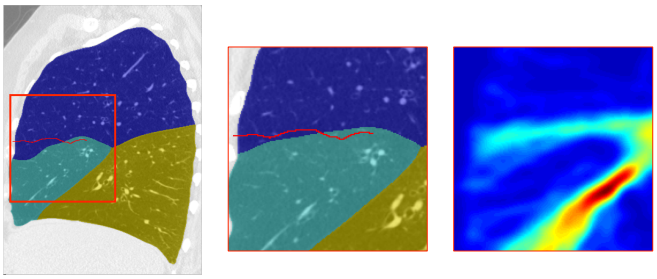
I have presented a novel lobe segmentation algorithm based on an unsupervised segmentation of the fissures with iterative false-positive removal, the creation of a groupwise fissure prior and a cost function combining patient and population information. Our algorithm does not require prior training or manual labelling to segment the fissures and build a population prior of the fissures. Fissure probabilities were obtained by parameterising a fissure enhancement filter with a Gaussian Mixture Model (GMM). Smoothness and curvature constraints were enforced in the segmentation by considering a Markov Random Field (MRF) regularisation. This led to rejection of most false-positives leading to high maximum F_1-score of 0.90. A method to construct a groupwise fissure prior given complete and incomplete fissures in a population was presented. I evaluated its role in identifying incomplete fissures whilst minimising potential segmentation failures. The method was validated on 55 cases from the LOLA11 study [99] and on 80 datasets from the COPDGene study [167]. I illustrated its applicability in correctly segmenting the lobes of patients with varying levels of disease severity and fissure incompleteness.



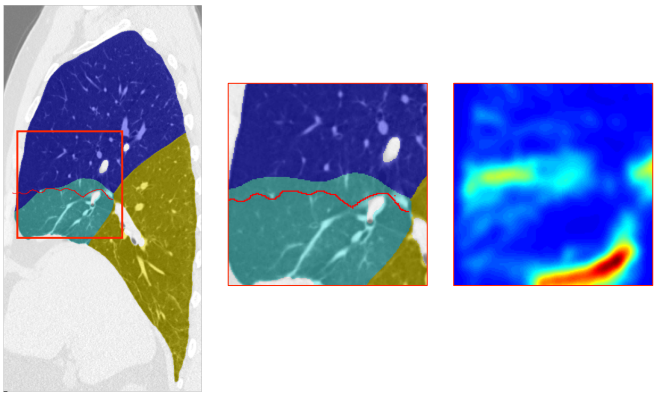
(a) Example 1 - correction in left major



(b) Example 2 - correction in right major



(c) Example 3 - correction in right minor



(d) Example 4 - correction in right minor

Figure 4.14: Lobe segmentation results without and with the groupwise prior (Π) in four different patients. The reference for non-visible fissures (label 2 and 3) is overlaid on the lobe segmentation. The colour of the reference was chosen to aid the visualisation and is not representative of the segmentation label. © 2017 IEEE. Reprinted, with permission from F. Bragman et al., Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior, IEEE Transactions on Medical Imaging, 2017.

Correct removal of false-positive fissures is necessary to accurately segment the fissures. Whilst supervised filter techniques have been utilised to segment the fissures [223, 118, 174], they require a training dataset to perform the classification. Manual labelling of voxels is laborious and may not always be practicable in a clinical setting. The fissure segmentation used by van Rikxoort et al. [221] required a training set as part of their algorithm and can only be employed when manual labelling is possible by an expert. Moreover, the applicability of a training set built on an independent set of scans applied to those acquired on different scanners is debatable.

Ross et al. [174] exploited a deformable model to identify the fissure surfaces, which may fail when the patient anatomy cannot be modelled by the atlas. The fissure enhancement applied by Lassen et al. [109] required experimentation to yield optimal ranges for the Hessian eigenvalues. Applying pre-existing thresholds to new datasets and those acquired at lower doses can be problematic and may cause undesirable drops in algorithm specificity. Applying hard-constraints on eigenvalue magnitudes in new scans may not be beneficial. The eigenvalue range may differ whilst the ratio is expected to remain constant. This is because the ratio will model different orientation patterns unique to various structures (spherical, tubular and sheet-like). This may cause an over-segmentation with too many false positives. The limitation of pre-existing threshold can also be applied to the filter of Wiemker et al. [238], who developed a weighting term based on the expected intensity of the fissures. It requires specific knowledge about the HU of the fissures to determine parameters unique to scans. Finding optimal parameters that yield a robust filter across a broad range of datasets is difficult and not desirable.

Our segmentation framework requires little prior knowledge regarding algorithm parameters and will be more robust than methods requiring prior training. Our fissure enhancement filter is based on ratios of Hessian eigenvalues. Given a set of new scans, separability between class distributions will exist since relationships between the eigenvalues should remain constant. Since we learn model parameters of the GMM, the segmentation of the fissures will be flexible when processing new scans. Classifying voxels by learning the underlying class distributions will capture a range of filter values; which can vary for each dataset. The integration of the MRF regularisation increases the robustness of the segmentation to noise. A voxel with a poor filter response can still receive a high fissure class probability when considering neighbourhood constraints on pairwise probabilities and

Hessian eigenvectors.

I showed in Section 4.4.1 that our method achieved a high median F_1-score of 0.90 and insensitivity to the filter input parameters. These parameters govern the separability of the class distributions and do not rely on knowledge of specific CT features. There is a minor dependence on the initialisation of the model parameters in the GMM. However, the iterative framework for increasing the MRF regularisation means no user-interaction is required. The iterative increase also deals with false-positive rejection as constraints on neighbourhood properties are given more weight until algorithm convergence. Performance of the fissure segmentation could be improved in future work by modelling the signal as a mixture of skew-normals akin to the work of Häme et al. [84] since I chose a GMM for mathematical simplicity.

I evaluated our algorithm on the LOLA11 dataset [99], which enabled direct comparison with the work of Lassen et al. [109] and van Rikxoort et al. [220]. I achieved the highest score of 0.884 in comparison to Lassen et al. [109] (0.881) and van Rikxoort et al. [220] (0.851). Both algorithms used superior lung segmentation algorithms (0.947 (our method) versus 0.962 [220] and 0.971 [109]), which may have had detrimental effects on our lobe segmentation scores in the most challenging cases (e.g. Figure 4.12g, h and j).

The quantitative experiment on dataset 2 highlighted the accuracy of the algorithm in areas with varying fissure visibility (Table 4.2). The high standard deviation associated with the segmentation of the right minor fissure was due to the failure of lobar segmentation in $2/30$ cases. In these two cases, post-processing of airway labelling errors could not be automatically corrected. The respective fissure means and standard deviations were: 1.52mm $\pm$ 1.49mm, 46.06mm $\pm$ 30.74mm and 1.33mm $\pm$ 2.06mm for the right major and minor fissures and and left major fissure respectively, which displays the isolated error. The large maximum errors in certain cases with low mean distances occurred in isolated areas close to the ribcage and near the lung hila where the automated segmentation disagreed significantly with the radiologist.

Qualitative testing (Table 4.3) highlighted the ability of the algorithm to produce good segmentation results in cases with higher severities of disease. The low proportion of lobar boundaries scoring 5 (errors $<$ 3mm) in cases with complete fissures (Table 4.3a) are a result of the narrow boundary definitions of the scoring system. The scoring system may not adequately reflect the performance of the algorithm since an isolated error will reduce

the score to 4 when it would otherwise be graded as 5. Most of these errors were less than 6mm from the reference fissure line and occurred in isolated regions prone to artefacts such as close to the rib-cage, the lung hila and the intersection between the right major and minor boundaries. Importantly, the algorithm was able to interpolate incomplete fissures (Table 4.3b) and demonstrated equivalence in performance to cases with complete fissures. The findings highlight the ability of the groupwise prior in conjunction with information from the vessel and airway tree to successfully guide fissure segmentation towards correct locations as defined by the reference standard.

My work bears many similarities with the implementation of Lassen et al. [109]. They also exploited information from auxiliary structures by combining the airway and vessel tree with the segmented fissures to create a cost image for watershed segmentation. Our algorithm differs primarily in the fissure segmentation and in the inclusion of population information in the cost image. Priors dependent on the segmentation of the vessel and airway tree might not always be fully informative. Airway tree segmentation is challenging and may not be sufficiently segmented to provide enough information about the location of lobar borders. The vessel tree may also not provide sufficient information in areas of largely incomplete fissures. The additional information provided by the groupwise fissure prior helps mitigate these issues. Within the LOLA11 dataset, the effect is marginal on the overall scores (0.884 versus 0.881). The dataset included many highly irregular scans, which made it difficult to create an accurate population model. The effect of the groupwise prior was more noticeable in dataset 2 from the COPDGene study (Figure 4.6 and 4.14) where I demonstrated the utility of the groupwise fissure.

The technique by which I construct the groupwise prior has the advantage of not requiring any pre-existing data. This does not require an expert to manually delineate complete fissures including visible and non-visible fissures. Since I construct the prior on the current set of data, it is not biased towards particular types of imaging data. The groupwise prior acted as a guide or region of confidence within the patient space rather than rigidly guiding the segmentation based on shapes in the training set, facilitating the segmentation of lobes of varying shape. Its effect was demonstrated on dataset 2 (Table 4.4) and visualised in Figure 4.14. Our method for constructing the groupwise prior suffers from a lack of flexibility in comparison to deformable models. A prior created by averaging all segmented fissures in the groupwise space may be over simplistic. Despite the simplicity of its construction,

the application can bias the results in certain cases. This was seen in the relationships calculated between segmentation errors and fissure incompleteness (Section 4.4.2.2) but also when quantifying the effect of the prior (Section 4.4.3). Since the average fissure is directly added within the patient space, it does not take into account the shape of the segmented fissure. This introduces a bias if the patient anatomy differs significantly from the mean. This occurred mostly in the right minor fissure, where several modes in the population exist (Figure 4.5). This led to a smaller average increase in the errors when testing algorithm performance without the prior (Table 4.4).

The dependence of our work on the construction of a groupwise space is a limitation. Groupwise registration is computationally expensive and measuring registration accuracy of inter-patient registration is difficult. Errors in the registration may be present, which can decrease the strength of our calculated groupwise prior. The work of Li et al. [112] used annotated landmarks from the airway tree to drive inter-patient landmark and intensity-based registration, indicating the applicability of detecting landmarks in inter-patient registration. Inclusion of the vessel density map in a multi-modal registration scheme to construct our groupwise space may then produce a more accurate fissure prior. Improving the inter-patient registration is then likely to increase the flexibility of the fissure prior. The transformation between the average space and the patient space acts as a deformable model. A more accurate mapping; obtained by including extra morphological information will help deform the groupwise prior to more unusual geometries.

The need to segment the airway tree may also decrease the applicability of our framework. Segmentation of the bronchial tree is an important determinant in the success of our algorithm as demonstrated by several failures in dataset 2. The vessel density map produces a good approximation of the fissure location, which may negate the need to use the airway tree in the fissure segmentation and cost function. However, initial seed labelling for the watershed still relies on the airway tree. Whilst I developed post-processing methods to make the method more robust to the quality of the airway segmentation, errors can lead to failures. A combination of the groupwise framework and the labelling method described in Section 4.2.4.3 may help generate better seeds without needing the airway tree.

The segmentation of incomplete fissures remains one of the biggest challenges in lobe segmentation. I presented the simultaneous construction of a groupwise prior to address this challenge. When the fissures are not-visible on CT, this is because they may be congenitally

absent or destroyed by inflammatory disease processes. The segmentation may therefore be creating an artificial division between lobes. In reality, the anatomical boundary between the lobes has either been destroyed or is absent. This is seldom mentioned in the lobe segmentation literature. When comparing our results with the label 3 reference, it is therefore important to note I am comparing algorithm extrapolation with the educated guess of an expert. Furthermore, it is not yet known what accuracy is needed in the segmentation of incomplete fissures to produce regional markers of disease that are clinically useful.

Despite the ability to correctly guide the segmentation in regions of incomplete fissures in most cases, the application of the groupwise prior requires further work. There may be issues when the mean of the population deviates significantly from the patient being segmented. In order to fully exploit the power obtained by fusing complete and incomplete segmentations, it is necessary to dynamically weight the groupwise prior in regions when it is needed and regions where information stemming from the patient is sufficient. Another solution may lie in creating various fissure models using different sets of patients from the population to mimic multi-atlas selection.

In conclusion, I have presented a lobe segmentation algorithm, which requires no prior training or manual labelling to both segment the fissures and build a population prior of the fissures. I have tested the method on 135 different datasets with varying levels of disease severity and complexity. The presented algorithm can be used in large studies to perform accurate regional quantification of disease progression and shows great promise to be integrated within a clinical setting.

Chapter 5

Manifold learning of COPD

The work presented in this chapter is based on and reproducted, with permission from F. Bragman, J. McClelland, J. Jacob, J. Hurst and D. Hawkes (2017)*. "Manifold Learning of COPD". In: *Medical Image Computing and Computer Assisted Interventions - MICCAI 2017*. Edited by M. Descoteaux, L. Maier-Hein, A. Franz, P. Collins and S. Duchesne. Lecture Notes in Computer Science, vol 10435, pp. 586-593.

∗ **Contributions**

I conceived the method and developed the framework presented in the paper. I performed all the data processing. I designed and undertook all the experiments, analysed the data and wrote the paper. Jamie McClelland and David Hawkes helped prepare the paper. John Hurst and Joseph Jacob provided clinical input during the development of the methodology.

Abstract

In this chapter, I build on the framework proposed in Chapter 3 to better quantify the pathology and pathophysiology of COPD. The aim was to develop a method that better quantifies the spread of disease throughout the lung and captures its effect on lung deformation. Current methods for quantifying disease extent in the lung rely on global averages and lobar averages when lobe segmentations are possible. By extending the local sampling framework of Chapter 3 to measure the spread of emphysema, functional small airways disease (fSAD) and variations in local volume change, a simple and clinically intuitive method for phenotyping COPD from CT could be developed.

I present local disease and deformation distributions in this chapter. The disease distribution aims to quantify two aspects of parenchymal damage: locally diffuse/dense disease and global homogeneity/heterogeneity. The deformation distribution aims to link parenchymal damage to local volume change. These distributions can be exploited in various ways to study COPD. The distributions were exploited to quantify accurate inter-patient differences in large cohort of patients from COPDGene. A framework originally applied to study neonatal brain development was implemented to investigate the progression of the distributions. Manifold learning was applied to seek a low-dimensional embedding based on the distributions whilst fusion of the learned embeddings was employed to combine distinct aspects of COPD into a single model. I demonstrated the utility of the framework by comparing associations between learned embeddings and measures of severity. I have also illustrated the potential to identify trajectories of disease progression in a manifold space of COPD.

5.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a complex disorder arising from various pathological processes including emphysema and functional small airways disease (fSAD). The extent of emphysema and fSAD that make up overall disease burden in an individual patient can vary, which can affect lung physiology and the clinical manifestations of a patient [82, 104, 162, 206, 157, 180]. Both components of COPD can progress at different rates and can complicate prognostication. Optimising the quantification of disease extent in COPD may improve the precision of disease staging and monitoring.

Analysis of lung disease from Computed Tomography (CT) has typically relied on the use of global averages. Such metrics cannot capture the anatomical distribution of disease. Several methods have attempted to capture regional manifestations with varying results in the literature. Tanabe et al. [205] employed the fractal exponent D to quantify the size distribution of emphysema clusters. They showed through model simulations that a spatially heterogeneous progression of emphysema may occur as a result of smoking [205]. They further showed in a separate study that a more homogeneous distribution of emphysema contributes more to FEV_1 decline. In a similar longitudinal study, focusing on α_1-antitrypsin deficiency, Parr et al. [157] showed that basal emphysema was associated with stronger progression in gas exchange impairment yet that apical emphysema associated better with FEV_1 decline.

Heterogeneity between the upper and lower lobes was assessed by Ju et al. [104] who proposed a heterogeneity index (HI) to quantify apical and basal emphysema subtypes. They found that patients with apical emphysema in contrast to Parr et al. [157] had significantly better lung function. Gietema et al. [75] found a negative correlation between pulmonary function and apical predominant emphysema although they note that it may not be statistically strong enough to be of clinical utility. Although not applied to COPD, a first attempt at quantifying intra and inter-lobar differences in emphysema was performed by Yilmaz et al. [246]. They quantified lobar fractional tissue volume and found associations between inter and intra-lobar distributions of emphysema and pulmonary function.

These methods are all attempts at better measuring COPD. However, considering the complex nature of the lung anatomy and its variation in the population, quantifying inter-patient differences with these metrics may not be accurate enough. In contrast to neuroimaging where groupwise studies are plentiful [212], with methods such as voxel-based

morphometry [130], population-wide studies of COPD and other lung diseases have largely been limited to using these simple metrics that may not be fully illustrative of the disease process.

The work of Feragen et al. [65] addressed the complexity of the airway tree by introducing a family of kernels for comparing anatomical trees endowed with vector attributes. These were exploited to statistically test that COPD diagnosed airway trees come from different distributions than healthy airway trees. Importantly, they showed that COPD classification improved when using airway wall area percentage kernels instead of traditional methods that compute an average across airway branch generations [83]. The need for a more complex quantification of the lung anatomy is seen in classification methods that exploit regions of interest (ROI) measured throughout the lung [42, 77, 192, 193]. Sørensen et al. [192] developed a new distance metric based on bipartite-graph matching between two collections of ROIs, which outperformed classification using average emphysema. This is corroborated in the work of Cheplygina et al. [42] who approached COPD classification as a Multiple Instance Learning (MIL) problem. Within MIL, the goal is to build a classifier for a collection of feature vectors, referred to as instances. This idea extends naturally to COPD, where a patient may experience positive instances of disease (emphysematous regions), and false positive instances (healthy regions).

In this chapter, I present a new method that quantifies the spread of disease throughout the lung and, which goes beyond the global extent of disease to quantify various aspects of pathological tissues destruction. It can be applied to any imaging features such as lung deformation to quantify links between structural damage and lung deformation. The framework can be exploited to quantify inter-patient differences in a large patient cohort to examine disease progression in a cross-sectional cohort. This bypasses the need to construct a groupwise space to compare lung texture. It is based on locally quantifying tissue classification results and lung deformation to capture heterogeneity or homogeneity across the lung. The outcome of the method is a distribution that quantifies various aspects of lung pathophysiology that can be modelled with manifold learning to test associations with clinical hypotheses. This work builds on the framework presented in Chapter 3. The lung is composed of a collection of regions with different pathological and healthy characteristics. The distribution of ROIs is unique to the patient, represents a signature of the disease state and may potentially be used as a marker of disease progression.

5.2 Methods

5.2.1 Lung deformation and tissue classification

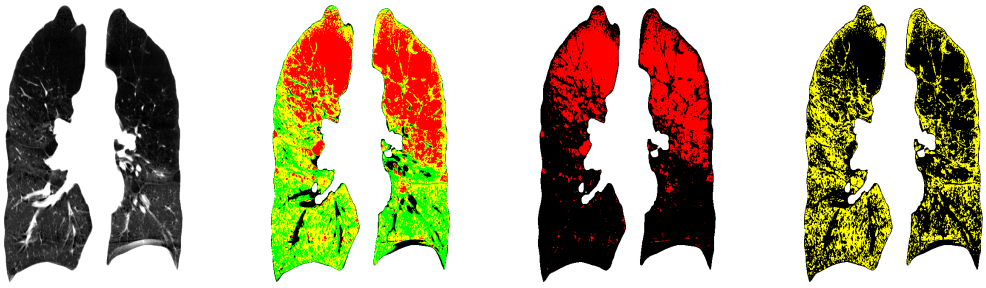
The deformation between paired breath-hold CT scans acquired at forced residual capacity ($\mathcal{I}_{exp}$, Ω^*) and total lung capacity ($\mathcal{I}_{ins}$, Ω) can be obtained using nonrigid registration. The output is a transformation $\boldsymbol{\varphi}$ mapping each coordinate $x \in \Omega \rightarrow x^* \in \Omega^*$. Local volume change on a voxel-wise basis is measured using the determinant of the Jacobian matrix: $J = \det(\nabla_x \boldsymbol{\varphi})$. A Jacobian determinant equal to 1 represents no volume change, compressive change occurs at $J < 1$ whilst $J > 1$ represents expansion.

Parametric Response Mapping (PRM) [70] is used to classify lung tissue into emphysema (PRM_{emph}) and functional small airways disease (PRM_{fSAD}). It is a simple technique that requires no training or manual labelling for the classification, which makes it particularly attractive in the studies with large cohorts of patients. The value of PRM_{fSAD} is not a direct measurement of functional small airways disease. The smaller airways cannot be directly measured since their dimensions fall below the resolution of modern CT scanners. It is an indirect measure as it classifies areas of gas trapping in $\mathcal{I}_{exp}$, which are not emphysematous in $\mathcal{I}_{ins}$.

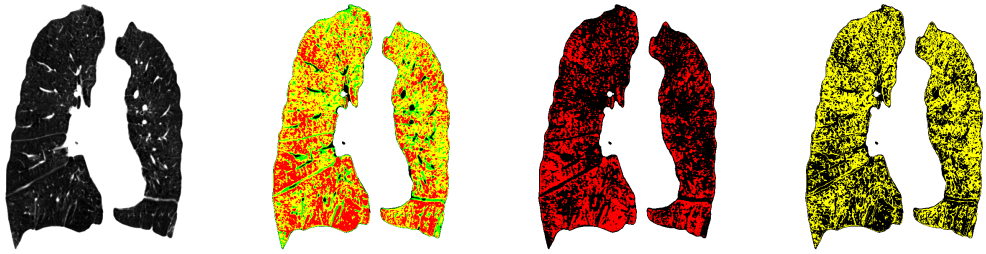
To perform the classification, PRM requires the transformation ($\boldsymbol{\varphi}$) resulting from the registration of $\mathcal{I}_{exp}$ to $\mathcal{I}_{ins}$. For all voxels $x_i \in \mathcal{I}_{ins}$, the tissue class z_i is obtained by applying the following Hounsfield Unit (HU) thresholds in $\mathcal{I}_{ins}$ and the registered $\mathcal{I}_{exp}$:

$$z_i = \begin{cases} \text{PRM}_{emph} & \text{if } \mathcal{I}_{ins}(x_i) \leq -950 \text{ and } \mathcal{I}_{exp}(\boldsymbol{\varphi}(x_i)) \leq -856 \\ \text{PRM}_{fSAD} & \text{if } \mathcal{I}_{ins}(x_i) > -950 \text{ and } \mathcal{I}_{exp}(\boldsymbol{\varphi}(x_i)) \leq -856 \end{cases} \tag{5.1}$$

The contribution of the airway walls and the vasculature is minimised by only considering voxels with an HU between -500HU and -1024HU in both scans. Equation 5.1 is applied to all voxels $x_i \in \mathcal{I}_{ins}$ to yield the classified volume $\mathcal{Z}$ (Figure 5.1).



(a) $\mathcal{I}_{ins}$ **(b)** $\mathcal{Z}$ **(c)** $\mathcal{Z} = \text{PRM}_{emph}$ **(d)** $\mathcal{Z} = \text{PRM}_{fSAD}$



(e) $\mathcal{I}_{ins}$ **(f)** $\mathcal{Z}$ **(g)** $\mathcal{Z} = \text{PRM}_{emph}$ **(h)** $\mathcal{Z} = \text{PRM}_{fSAD}$

Figure 5.1: Example of Parametric Response Mapping classification. The map for voxels classified as normal (PRM_{normal}), which are green voxels in $\mathcal{Z}$, is not displayed.

5.2.2 Local disease and deformation distributions

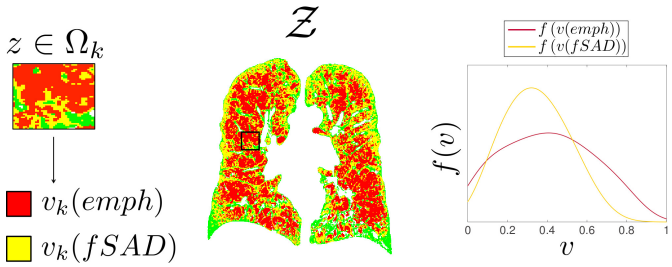
I present the concept of local feature distributions (Figure 5.2 and 5.3). The aim is to quantify local abnormalities in lung physiology and pathology to define a signature unique to a patients disease state. I introduce two models: 1) local disease distributions (Figure 5.2) and 2) local deformation distributions (Figure 5.3). The disease distributions model the spread of emphysema and fSAD whilst the deformation distribution characterises local volume change across the lung. They are created by locally sampling regions of $\mathcal{Z}$ and J in a Cartesian grid using local regions of interest Ω_k (ROI) where $k = 1 \cdots K$ indexes the center voxel of the ROI. The size $(r \times r \times r)$ of the ROI governs the scale of the sampling.

The local disease distributions model two properties of disease spread: 1) locally diffuse/dense disease and 2) global homogeneity/heterogeneity. For each ROI centered at z_k where $z \in \Omega_k$, I computed the fraction of PRM_{emph} and PRM_{fSAD} voxels; defined as $v_k(emph)$ and $v_k(fSAD)$. Dense disease occurred when $v_k(\cdot) \to 1$ whilst diffuse disease was present when $v_k(\cdot) \to 0$. The deviation of diffuse and dense regions in the lung defined the heterogeneity/homogeneity of disease spread.

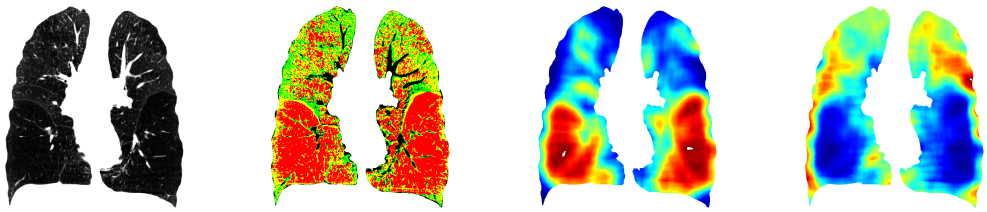
A distribution $f(v(\cdot))$ for each feature was built by sampling K regions. The shape of

the distribution is governed by the two disease properties (Figure 5.4). It provides information on the nature of local disease spread (diffuse or dense), whether it is homogeneous or heterogeneous. Additionally, it also captures whether the distribution is bimodal.

Expansion of the lung is dependent on local biomechanical properties (emphysema) and airway resistance (functional small airways disease). Both these processes will affect lung deformation locally. The local deformation distribution captures volume change on a local basis whilst modelling potential homogeneities or heterogeneities (Figure 5.3). To measure volume change on a local basis, the Jacobian map (J) was sampled by calculating the mean Jacobian ($\mu(J)_k$) for all Ω_k. A distribution $f(\mu(J))$ of these measurements was built to capture local volume change throughout the lung using the same process as for the disease distribution.



(a) Local sampling of $\mathcal{Z}$



(b) $\mathcal{I}_{ins}$ **(c)** $\mathcal{Z}$ **(d)** $v(\text{PRM}_{emph})$ **(e)** $v(\text{PRM}_{fSAD})$

Figure 5.2: Local disease distribution quantification. © 2017 Springer Nature. Reprinted, with permission from F. Bragman et al., Manifold learning of COPD, Medical Image Computing and Computer Assisted Interventions, 2017.

5.2.3 Manifold Learning of COPD distributions

The goal of manifold learning lies in learning a low-dimensional representation to model variability in the data. In this chapter, I have presented the concept of local disease and local deformation distributions. It is possible that one aspect in the heterogeneity of COPD is due to the different mechanisms of emphysema and fSAD and its effect on local lung biomechanics. The presented distributions were developed to capture these processes. I used manifold learning to model variations of the distributions in a large population of

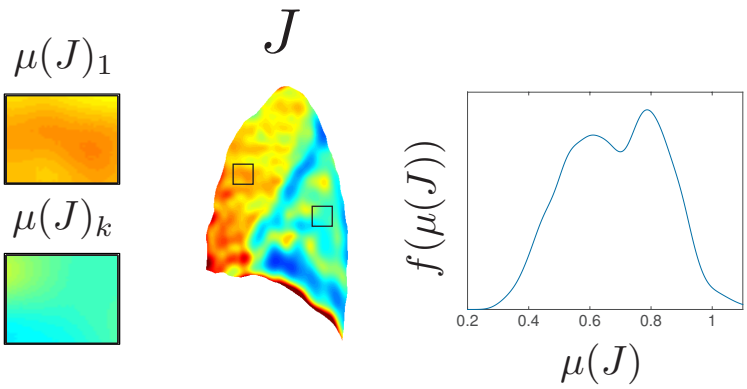
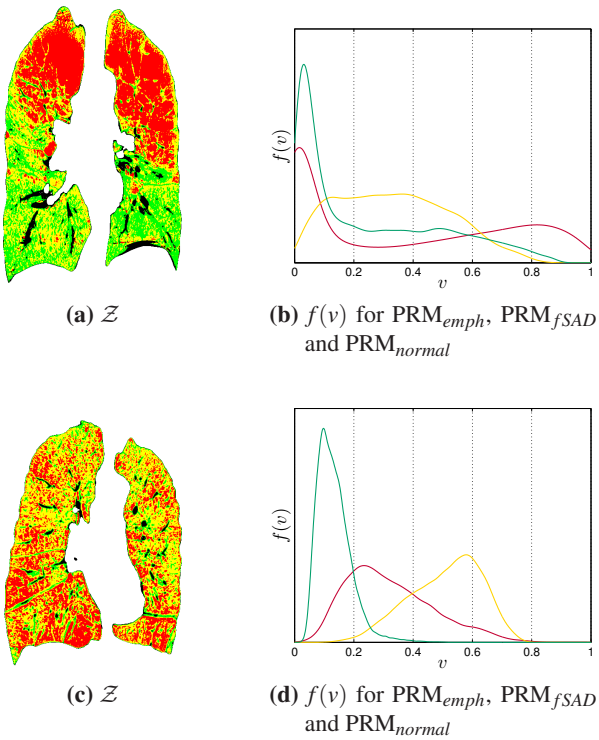


Figure 5.3: Local deformation distribution quantification. © 2017 Springer Nature. Reprinted, with permission from F. Bragman et al., Manifold learning of COPD, Medical Image Computing and Computer Assisted Interventions, 2017.



(a) $\mathcal{Z}$

(b) $f(v)$ for PRM$_{emph}$, PRM$_{fSAD}$ and PRM$_{normal}$

(c) $\mathcal{Z}$

(d) $f(v)$ for PRM$_{emph}$, PRM$_{fSAD}$ and PRM$_{normal}$

Figure 5.4: Local disease distributions for patient with equal levels of emphysema in the lung (μ(PRM$_{emph}$=30%). Top row patient: FEV$_1$%predicted=28.3 and bottom row patient: FEV$_1$%predicted=33.8

COPD patients. This led to separate embeddings for emphysema, fSAD and volume change. Fusion of these embeddings was then performed to create a single model of these processes to analyse COPD progression.

5.2.3.1 Distribution distances

Inter-patient differences can be computed by considering the distance ($\mathcal{L}$) between two feature distributions ($f_1(v)$ and $f_2(v)$). As the distance metric employed is an important determinant in comparing distributions, I considered the following: 1) the $\mathcal{L}_1$-norm, 2) the $\mathcal{L}_2$-norm, 3) the Earth-Movers Distance ($\mathcal{L}_{EMD}$) and 4) the Quadratic-Chi histogram dissimilarity ($\mathcal{L}_{QC}$). The first two are bin-to-bin distance measures and the latter are cross-bin metrics. The local distributions ($f_1(v)$ and $f_2(v)$) are quantised into histograms ($H(v), K(v) \in \mathbb{R}^{N_b}$) with N_b bins. These histograms are normalised such that they sum to one. They have equal mass and are probability mass functions on equal intervals ($[0,1]$). The probability mass at a bin i for the histograms H and K are defined as h_i and k_i.

$\mathcal{L}_p$ Distances The $\mathcal{L}_1$ and $\mathcal{L}_2$-norm are defined as follows:

$$\mathcal{L}_p(H,K) = \left(\sum_i^{N_b} |h_i - k_i|^p \right)^{1/p} \tag{5.2}$$

with $p = 1$ and $p = 2$ corresponding respectively to the $\mathcal{L}_1$ and $\mathcal{L}_2$-norm. They are considered bin-by-bin measures as the distances are computed by matching bins at the same index. A drawback of $\mathcal{L}_p$-style distances is their dependence on the bin number. If the number of bins is low, the resulting histogram will not be discriminative. If the number of bins is too high, quantisation of the data into bins with small widths can yield undesirable discontinuities in the histogram leading to a non-robust distance.

Earth-Movers Distance The Earth-Movers Distance (L_{EMD}) [175] is a metric based on the principle of optimal transport. It measures the minimum amount of work needed to transform one distribution into another. The EMD operates on signatures of histograms $s_j = \{m_j, w_j\}$, which represents a set of feature clusters. Each cluster is represented by its mean m_j and by the fraction of w_j of voxels that belong to that cluster. The Earth-Movers distance is defined as

$$\mathcal{L}_{EMD}(H,K) = \frac{\min\limits_{\{f_{i,j}\}} \left(\sum_{i,j} f_{i,j} d_{i,j} \right)}{\sum_{i,j} f_{i,j}} \tag{5.3}$$

where $d_{i,j}$ is the ground distance and $f_{i,j}$ is the flow between histogram bins i and j. The aim of $\mathcal{L}_{EMD}$ is to find the minimal flow from H to K. Importantly, if one assumes that the histograms are one-dimensional, have equal mass and have equal number of bins, the EMD has a closed-form solution equivalent to the $\mathcal{L}_1$-norm on the cumulative distribution function of $P(H)$ and $P(K)$ [110] such that

$$\mathcal{L}_{EMD}(H,K) = \mathcal{L}_1(P(H),P(K)) \tag{5.4}$$

Quadratic-Chi Distance The Quadratic-Chi distance $\mathcal{L}_{QC}$ is another example of a cross-bin distance measure [160]. It combines the Quadratic-Form distance ($\mathcal{L}_{QF}$) and the Chi-squared distance ($\mathcal{L}_{\chi^2}$) such that

$$\mathcal{L}_{QF}(H,K) = \sqrt{(H-K)^T A (H-K)} \tag{5.5a}$$

$$\mathcal{L}_{\chi^2}(H,K) = \frac{1}{2}\sum_i \frac{(h_i-k_i)^2}{(h_i+k_i)} \tag{5.5b}$$

$$\mathcal{L}_{QC}(H,K) = \sqrt{\sum_{i,j}\left(\frac{(h_i-k_i)}{(\sum_c (h_c+k_c)A_{c,i})^m}\right)\left(\frac{(h_j-k_j)}{(\sum_c (h_c+k_c)A_{c,j})^m}\right)A_{i,j}} \tag{5.5c}$$

where A is a bin-to-bin similarity matrix and m is a normalisation factor. The authors [160] recommend $m = 0.9$ for best results. In the context of the Quadratic-Chi distance, the similarity matrix is defined as:

$$A_{i,j} = 1 - \frac{D_{i,j}}{\max_{i,j}(D_{i,j})} \tag{5.6}$$

where D is a distance matrix between histogram bins.

The Quadratic-Chi distance exploits advantages of both $\mathcal{L}_{QF}$ and $\mathcal{L}_{\chi^2}$. The Quadratic-form distance takes cross-bin relationships through the similarity matrix A whilst the χ^2 distance reduces the effects of bin-to-bin distances caused by bins with large values.

5.2.3.2 Manifold Learning and fusion

Manifold learning is used to model emphysema, fSAD and Jacobian distributions. The aim is to capture variations in the distributions in a population of COPD patients. The

progression of emphysema and fSAD occur synchronously yet the pathological processes are distinct. Their combination is likely to affect the lungs ability to deform. The aim is to capture both processes and their effect on lung deformation to analyse COPD. The manifold fusion framework presented by Aljabar et al. [4] is employed to create a single representation of these processes (Figure 5.5).

For P subjects, the classified volumes $\mathcal{Z}_1, \cdots, \mathcal{Z}_P$ and their respective Jacobian determinant maps $J = J_1, \cdots, J_P$ are obtained. The local disease and deformation distributions are created and quantised using N_b bins into their respective histograms $h_{p,v(emph)}$, $h_{p,v(fSAD)}$ and $h_{p,J}$. Pairwise measures in the population are obtained using any of the distance functions $\mathcal{L}$ discussed above in Section 5.2.3.1. This yields the pairwise matrices $\mathcal{M}^{emph}$, $\mathcal{M}^{fSAD}$ and $\mathcal{M}^{J}$ for emphysema, fSAD and the Jacobian respectively.

The pairwise matrices can be visualised as connected graphs where each node represents a patient and the edge length between nodes is the distance $\mathcal{L}$. Various linear and non-linear dimensionality reduction methods can be applied to the matrices $\mathcal{M}^{(\cdot)}$. Since the edge-weights in $\mathcal{M}^{(\cdot)}$ are distances, the Isomap algorithm [209] is a natural choice as it interprets edge-weights in $\mathcal{M}^{(\cdot)}$ as distances.

Th goal of Isomap is to find the mapping function $f : \mathcal{M}^{(\cdot)} \rightarrow \mathbb{R}^d$, where $\mathcal{M}^{(\cdot)} \in \mathbb{R}^D$ such that $\mathcal{M}^{(\cdot)}$ is projected to a low-dimensional space $\mathbb{R}^d$. The main assumption of Isomap is that f is an isometric chart, which means the distances are preserved as they are mapped to the new space. If x_i and x_j are points on the manifold and $\mathcal{D}(x_i, x_j)$ is the geodesic distance between them, the following relationship is obeyed:

$$\left\| f(x_i) - f(x_j) \right\| = \mathcal{D}(x_i, x_j) \tag{5.7}$$

A further assumption of the Isomap algorithm is that the geodesic distance between nearby points is linear and can be approximated by the Euclidean distance. For points which cannot be approximated by a linear function, the Euclidean distance will not be a valid approximation. A different approach is therefore needed to compute the distance between these points. The Isomap algorithm first constructs a sparse representation of $\mathcal{M}^{(\cdot)}$ by searching for the K-nearest neighbours (k-NN) weighted by Euclidean distances between each node. Local structure can thus be modelled by an appropriate choice of the parameter k. To estimate the manifold structure, a full pairwise geodesic matrix $\mathcal{D}$ is estimated by identifying the shortest path in the k-NN graph of $\mathcal{M}^{(\cdot)}$. This is performed

using Djikstra's shortest-path algorithm [51]. Given the matrix $\mathcal{D}^{(\cdot)}$, Isomap then seeks the lower-dimensional embedding coordinates $y_p^{(\cdot)}, p = 1, \cdots, P$ that satisfies Equation 5.7 by minimising the cost function:

$$\min \sum_{p,j} \left(\mathcal{D}_{p,j}^{(\cdot)} - \left\| y_p^{(\cdot)} - y_j^{(\cdot)} \right\| \right)^2 \tag{5.8}$$

using Multi-Dimensional Scaling. The Isomap algorithm is applied separately to $\mathcal{M}^{emph}$, $\mathcal{M}^{fSAD}$ and $\mathcal{M}^{J}$, which yields the respective embeddings y^e, y^f and y^J with dimension d^e, d^f and d^J that need to be selected (Figure 5.5).

The coordinates $y^{(\cdot)}$ can boe fused together in any combination to create various models and investigate different combinations of features implicated in COPD. For simplicity, I consider the fusion of all embeddings. The coordinates are uniformly scaled with scaling factors s^e, s^f and s^J such that the first component of each embedding $y_{p,1}^{(\cdot)}$ has a unit variance. These are concatenated to yield the new set $Y_p = (s^e y_p^e, s^f y_p^f, s^J y_p^J)$ with dimension $d^e + d^f + d^J$. A new distance matrix $\mathcal{D}^c$ is obtained by calculating the Euclidean distance between all pairs of coordinates in Y. A second Isomap step is applied to yield the combined coordinate embedding y^c with dimension d^c (Figure 5.5). The new embedding coordinates (y^c) can then be used in a classification framework to find clusters of patients with similar presentations or in regression to test associations with various clinical metrics and outcomes.

5.3 Experiments and results

5.3.1 Histogram quantisation and pairwise distances

There are various parameters that need to be set to run the framework. In the analysis of local disease and deformation distributions, quantisation of the distributions into probability mass functions is necessary to compute pairwise distances. Pairwise distances between patients provides a metric that quantifies potential differences in textural damage and deformation. The choice of distance measure will affect the accuracy of the measurements and any ensuing analysis. The choice of bin size will affect the shape of the mass function (Figure 5.6) and will affect the ensuing distances. Adaptive binning techniques exist [153], which can be approximated as $N_b = \sqrt[3]{N_s}$ where N_s is the number of sampled ROIs. However, N_s varies for each patient and an equal N_b is necessitated to use the closed-form solution of the $\mathcal{L}_{EMD}$ (Equation 5.4). It is noteworthy that the choice of distance function ($\mathcal{L}$) could be learnt in the training phase of the algorithm when considering classification

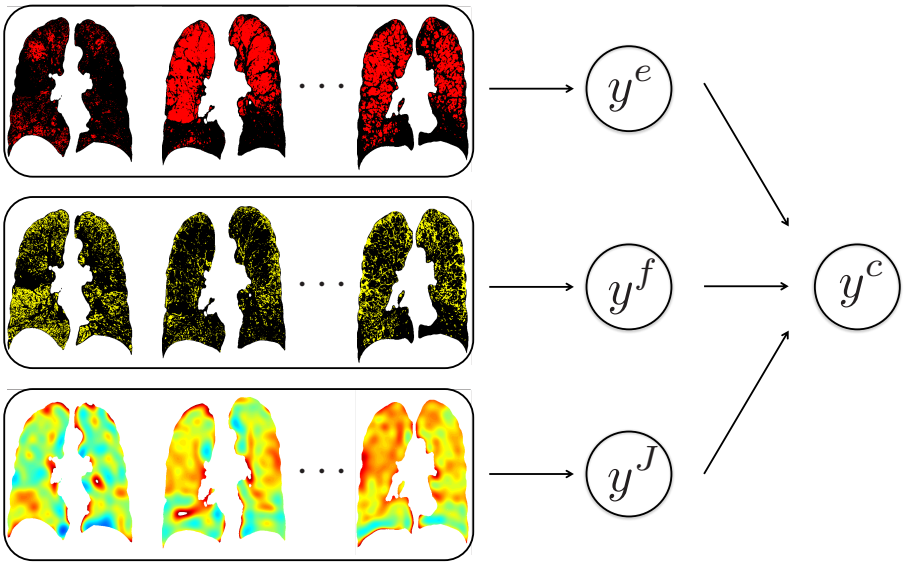
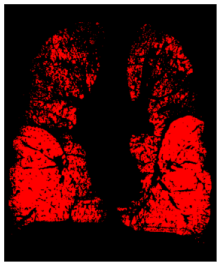
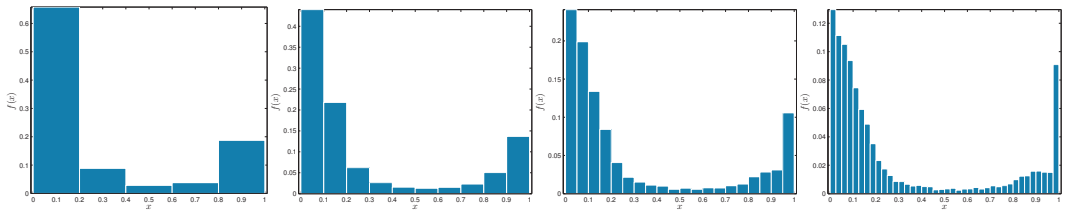


Figure 5.5: Overview of manifold learning and fusion of local disease and deformation distributions. Individual embeddings y^e, y^f and y^J are quantified for each set of distributions ($f(\cdot)$) obtained by sampling emphysema ($\mathcal{Z} = \text{PRM}_{emph}$), fSAD ($\mathcal{Z} = \text{PRM}_{fSAD}$) and Jacobian determinant maps (J). A final combined embedding (y^c) is obtained by fusion of the individual coordinates.

[193].



(a) $\mathcal{Z} = \text{PRM}_{emph}$



(b) $N_b = 5$ **(c)** $N_b = 10$ **(d)** $N_b = 20$ **(e)** $N_b = 40$

Figure 5.6: Effect of N_b on the distribution.

Within this experiment, I examined the robustness of pairwise distances on the four distance metrics ($\mathcal{L}_{1,2,EMD,QC}$, Section 5.2.3.1) with respect to the number of bins. The aim was to determine the distance measure that is most robust to quantisation, an approximation of the optimal number of bins and choose the distance measure, which best matched

perceptual differences.

5.3.1.1 Effect of N_b on the distance metric

I created local-disease distributions (emphysema and fSAD) for 198 COPD patients (GOLD ≥ 1) from the COPDGene cohort [167]. These were created using a local sampling neighbourhood size of $r = 20$mm. The distributions were created by a dense sampling of 5mm in the x, y and z directions. This led to an average $N_s = 44,251 \pm 11,572$ number of sampled regions. Histograms were created by iteratively increasing the number of bins by 10 at each iteration within the range: $N_b(i) \in [5,255]$. Pairwise matrices ($\mathcal{M}^e_{\mathcal{L}_{(\cdot)}}(i)$ and $\mathcal{M}^f_{\mathcal{L}_{(\cdot)}}(i)$) were created at each iteration $N_b(i)$ by considering the four proposed distance metrics.

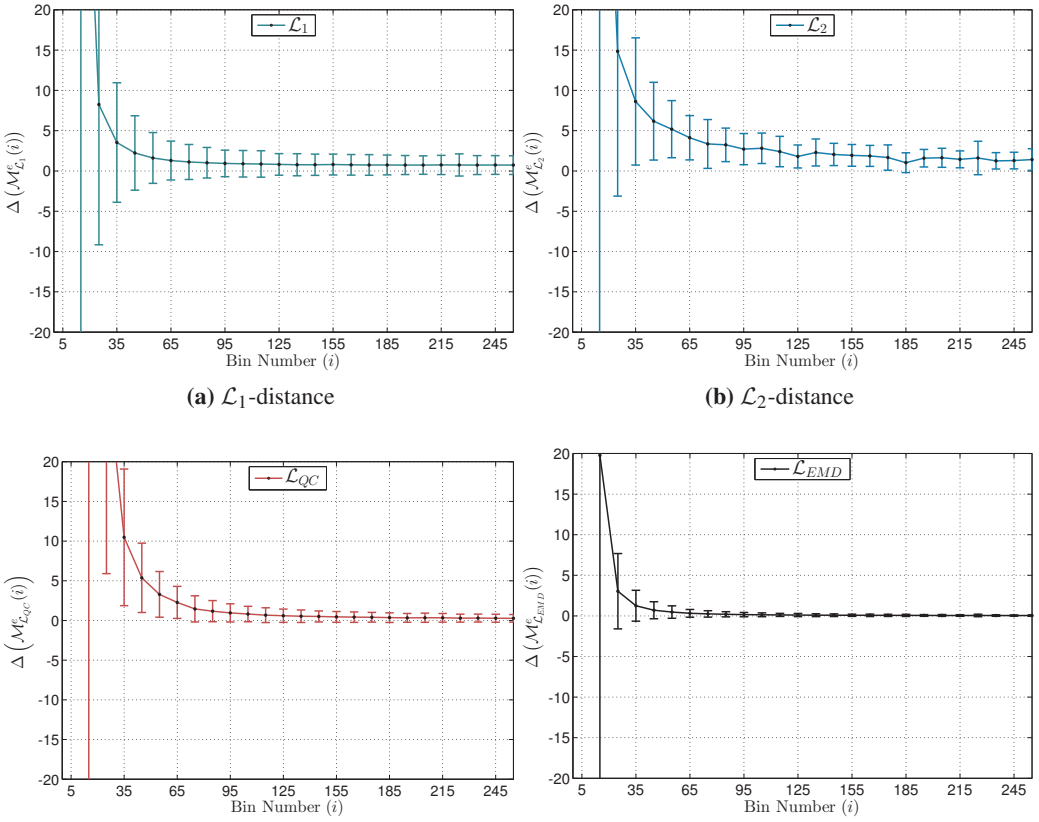
In order to compare how the pairwise distances change with bin size, I first normalised the pairwise distances by the maximum of all pairwise distances at each iteration. Since the pairwise matrices are symmetric, I only considered the upper triangular matrix $\mathcal{M} := U(\mathcal{M})$. The effect of N_b on the pairwise distances was then measured with the absolute percentage change of the normalised distances:

$$\Delta\left(\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i)\right) = \frac{\left|\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i) - \mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i-1)\right|}{\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i-1)} \tag{5.9}$$

where $\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i)$ and $\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i-1)$ are the pairwise distances at bins i and $i-1$ for any feature and choice of distance function. Assessing the relationship between $\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i)$ and $N_b(i)$ for each distance function provides an indication of the effect of N_b on pairwise measures in the cohort of patients.

For each bin iteration $N_b(i)$, I calculated the mean and standard deviation of $\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i)$ (Figures 5.7 and 5.8). I also computed the maximum change of $\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i)$ (Figure 5.9). The biggest change in the pairwise distances occurred at low bin numbers ($N_b \in [5,45]$). This is expected as the distribution is likely to vary significantly when increasing the number of bins from a very coarse distribution (Figure 5.6). At higher values of N_b (> 40), changes in histogram distances became less pronounced (Figures 5.7 and 5.8). However, I noticed a difference between the bin-to-bin ($\mathcal{L}_p$) and cross-bin ($\mathcal{L}_{EMD}$ and $\mathcal{L}_{QC}$) measures. In the former, despite convergence of the pairwise distances, there was still a large variability with increasing bin size. Conversely, the cross-bin measures saw greater robustness (in particular the $\mathcal{L}_{EMD}$) with regards to the bin number (when $N_b > 30$). This was specifically seen when assessing the maximum $\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}(i)$ with increasing N_b for each metric (Figure 5.9).

Increasing the bin size can cause large changes in the $\mathcal{L}_p$ distances. The $\mathcal{L}_{EMD}$ provided the best robustness with very low maximum percentage changes. This was expected in the $\mathcal{L}_{EMD}$ as the work needed to transform distributions H and K is unlikely to vary when an appropriate number of bins is used.



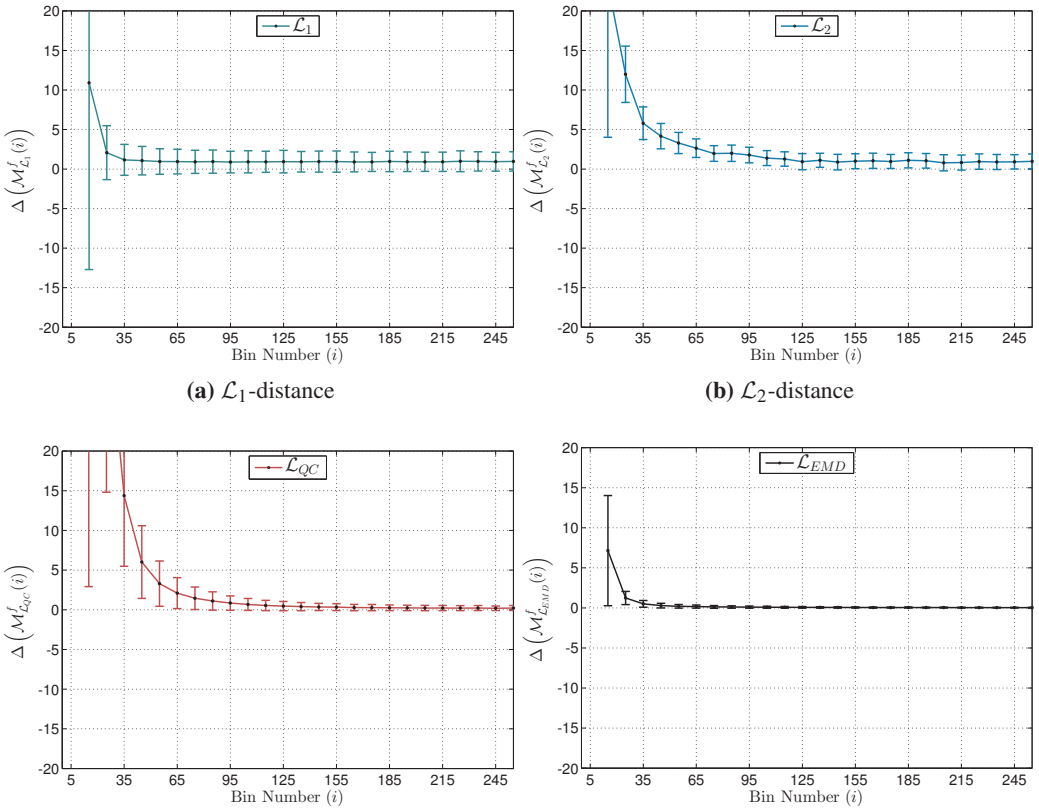
(a) $\mathcal{L}_1$-distance

(b) $\mathcal{L}_2$-distance

(c) $\mathcal{L}_{QC}$-distance

(d) $\mathcal{L}_{EMD}$-distance

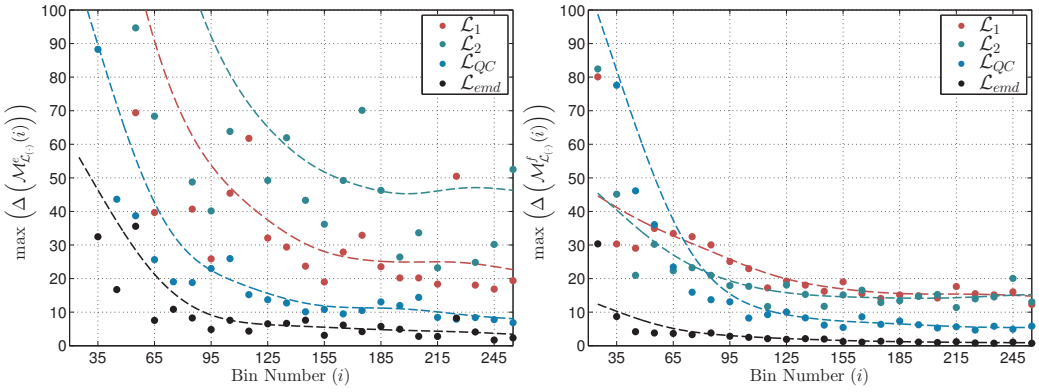
Figure 5.7: PRM$_{emph}$: $\Delta\left(\mathcal{M}^e_{\mathcal{L}_{(\cdot)}}\right)$ versus N_b.



(a) $\mathcal{L}_1$-distance **(b)** $\mathcal{L}_2$-distance

(c) $\mathcal{L}_{QC}$-distance **(d)** $\mathcal{L}_{EMD}$-distance

Figure 5.8: PRM$_{fSAD}$: $\Delta\left(\mathcal{M}^f_{\mathcal{L}_{(\cdot)}}\right)$ versus N_b.



(a) PRM$_{emph}$: $\mathcal{M}^e_{\mathcal{L}_{(\cdot)}}$ **(b)** PRM$_{fSAD}$: $\mathcal{M}^f_{\mathcal{L}_{(\cdot)}}$

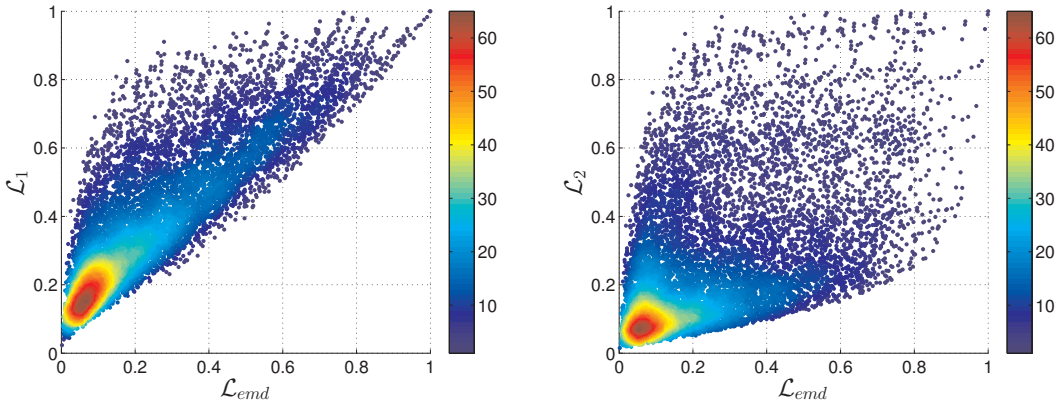
Figure 5.9: Maximum % $\Delta\left(\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}\right)$ versus N_b. The curves were fitted with Nadaraya-Watson kernel regression using a Gaussian kernel with a bandwidth of 25.

5.3.1.2 Perceptual differences in distributions

The previous experiment examined the effect of N_b on various histogram distances. The cross-bin measures exhibited more robustness to histogram quantisation. The $\mathcal{L}_{EMD}$ provided the best stability, as shown by the maximum change in $\Delta\left(\mathcal{M}^{(\cdot)}_{\mathcal{L}_{(\cdot)}}\right)$ with N_b increase (Figure 5.9). The $\mathcal{L}_p$-distances displayed significantly higher instability with respect to N_b.

This experiment builds on the previous by analysing specific cases where cross-bin distance metrics significantly differ from bin-to-bin metrics. This is to examine if the distances given by $\mathcal{L}_{(\cdot)}$ match the perceptual similarity of the classified volumes $\mathcal{Z}$. I chose the $\mathcal{L}_{EMD}$ as the reference for cross-bin distances and compared it to $\mathcal{L}_p$-distances. The distances were normalised to the maximum distance in the population. Analysis was performed at $N_b = 95$, which was chosen by judging convergence of all plots in Figures 5.7 and 5.8.

I first assessed relationships between $\mathcal{L}_{EMD}$-$\mathcal{L}_1$ and $\mathcal{L}_{EMD}$-$\mathcal{L}_2$ (Figure 5.10). I observed large non-linear relationships between the cross-bin and bin-to-bin distances. There were cases in both relationships where the $\mathcal{L}_p$ distances produced large distances ($\mathcal{L} > 0.6$) in comparison to a low $\mathcal{L}_{EMD}$ ($\mathcal{L} < 0.3$). There were also cases where $\mathcal{L}_{EMD}$ produces a large distance ($\mathcal{L} > 0.8$) yet the $\mathcal{L}_2$-distance was small ($\mathcal{L} < 0.3$).



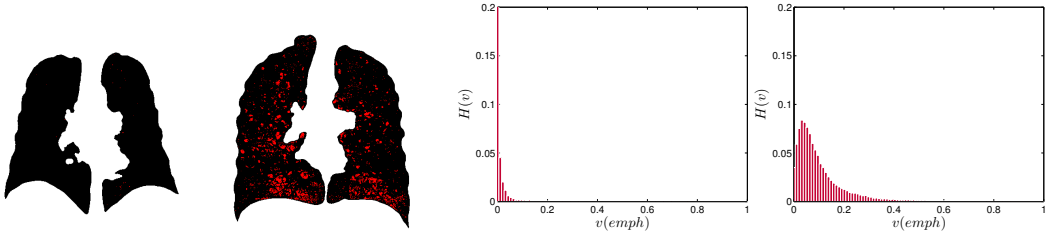
(a) $\mathcal{L}_{EMD}$ versus $\mathcal{L}_1$ **(b)** $\mathcal{L}_{EMD}$ versus $\mathcal{L}_2$

Figure 5.10: Relationship between the $\mathcal{L}_{EMD}$ and $\mathcal{L}_p$ distances. The density of the scatter plot is overlaid.

These cases were assessed to understand instabilities in the metrics and their utility in distinguishing differences in the distributions. In Figure 5.13, various pairs of patients with similar patterns of emphysema progression can be observed. The patterns can be differentiated by the type of distribution and can be either classified as exponential or skew-normal distributions. An exponential distribution characterises very early disease patterns

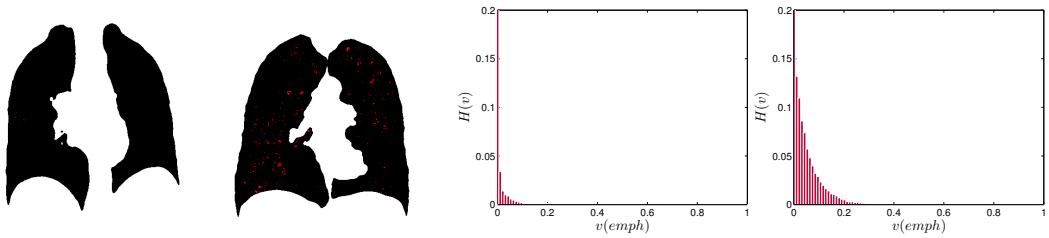
with minimal areas of local disease. A skew-normal distribution represents lung pathology at a later stages with various regions of diffuse disease and certain isolated areas with a more dense pattern. It is observable that there exists an instability in the bin-to-bin distance when comparing exponential and positive skew-normal distributions.

Due to the high density of regions with values close to 0 in the exponential family, an $\mathcal{L}_p$-type distance will yield large bin-to-bin distances when computing the difference with a skew-normal distribution. This will artificially inflate the pairwise distance. In contrast, the amount of work needed to transform an exponential into a skew-normal is low and translates to a small $\mathcal{L}_{EMD}$. This matches the perceptual differences between both classified volumes. Considering the robustness of the $\mathcal{L}_{EMD}$ towards the bin number and its stability with respect to these special cases, the $\mathcal{L}_{EMD}$ was employed in the proposed framework.



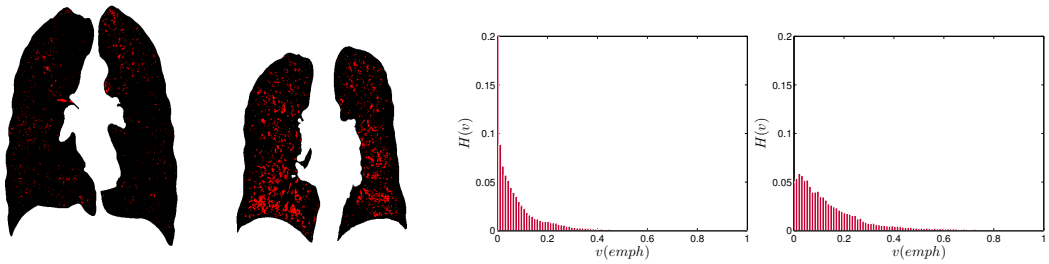
(a) Exponential **(b)** Skew-normal **(c)** Exponential

Figure 5.11: $\mathcal{L}_1 = 0.88$ and $\mathcal{L}_{EMD} = 0.18$



(a) Exponential **(b)** Exponential **(c)** Exponential **(d)** Exponential

Figure 5.12: $\mathcal{L}_1 = 0.69$ and $\mathcal{L}_{EMD} = 0.09$



(a) Exponential **(b)** Skew-normal **(c)** Exponential **(d)** Skew-normal

Figure 5.13: $\mathcal{L}_1 = 0.91$ and $\mathcal{L}_{EMD} = 0.28$

5.3.2 Manifold learning and fusion of disease and deformation distributions

5.3.2.1 Data

Patients with COPD (GOLD ≥ 1) from the COPDGene database [167] were analysed. A total of $1,154$ scans were downloaded from the COPDGene database. Scans from various scanners (GE Medical Systems, Siemens and Philips) were used. The STANDARD (GE), AS+ B31f and B31f (Siemens), and 64 B (Philips) reconstruction algorithms were employed.

5.3.2.2 Data processing

The Pulmonary Toolkit [52] was used with the algorithm of Hu et al. [100] to segment the lungs. Breath-hold scans were registered using NiftyReg [139]. A modified version of the Niftyreg EMPIRE10 pipeline [137] was used. The transformation was a stationary velocity field parameterised by a cubic B-spline and the similarity measure was MIND [91]. The constraint term was the bending energy of the velocity field and was weighted at 1% of the objective function for all stages of the pipeline. The transformation obtained from the registration was used to resample the expiration scan into the space of the inspiratory scan. Lung tissue was then classified using the PRM [71] by exploiting the coregistered scans.

To quantify the local deformation and disease distributions, I used a sampling size of $r = 20mm$ for the ROI. This is consistent with the size of the secondary pulmonary lobule. Sampling was performed by creating a Cartesian grid of center voxels spaced every 5mm. I chose a value of $N_b = 60$ as its effect on $\mathcal{L}_{EMD}$ pairwise distances was minimal with increasing N_b when N_b was greater than 50 (Section 5.3.1.1). The Earth Movers distance ($\mathcal{L}_{EMD}$) was used as the distance function to calculate the pairwise distances and create $\mathcal{M}^{(\cdot)}$ for each feature.

5.3.2.3 Quality control of the registration results

The local disease and deformation distributions are both dependent on the quality of the registration. Quality control was therefore performed to reject registrations that may affect the quality of the modelling. A 5-point qualitative scoring system was first used to assess the quality of the registrations (Table 5.1). A registration was rejected if the score was ≤ 2 and accepted otherwise. Qualitative scores for 300 subjects were recorded and a frequency distribution of the scores is displayed in Figure 5.14. The remaining patients from the COPDGene cohort were assessed using the scoring criteria and were marked as accept or reject. After manual inspection of the registration results and use of the scoring criteria,

Table 5.1: Qualitative scoring system for registration quality control

Score	Criteria
5	No noticeable errors.
4	Small isolated errors not likely to affect subsequent analysis.
3	Larger errors not likely to affect subsequent analysis.
2	Systematic larger errors that will affect subsequent analysis.
1	Large failures in the registration.

$P = 743$ patients were selected for the analysis. Examples of registration results for the defined quality control scores can be viewed in Figure 5.15.

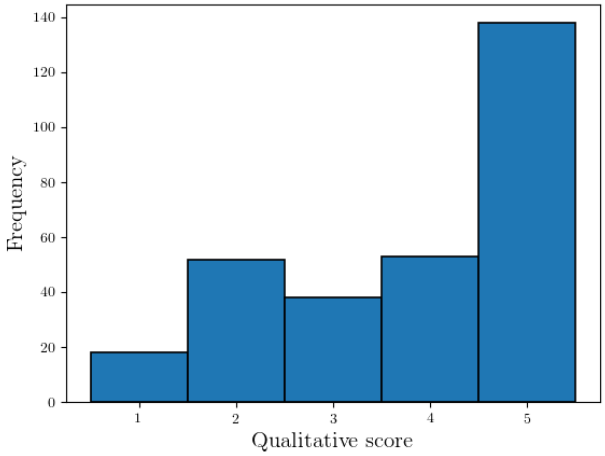


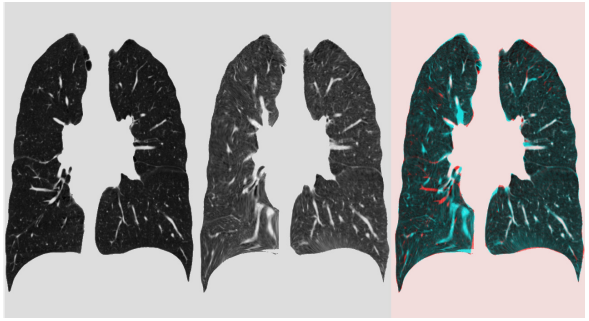
Figure 5.14: Frequency distribution of registration qualitative scores

5.3.2.4 Manifold learning with Isomap

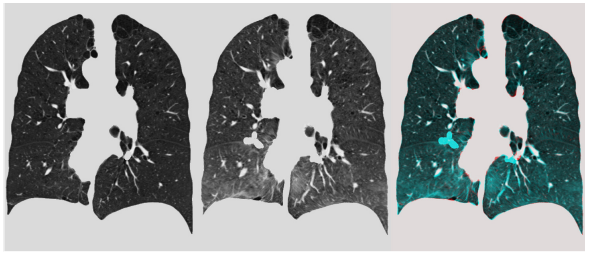
Manifold learning of each similarity matrix was performed with the Isomap algorithm [209] using the Dimensionality Reduction Toolbox [216]. The output of each Isomap step is a set of embedding coordinates y of dimensionality d. The dimensionality d and the local neighbourhood parameter K were determined for each embedding by estimating the reconstruction quality of the lower-dimensional coordinates. The residual variance defined as

$$\rho_{res} = 1 - \rho^2_{\mathcal{D},y} \tag{5.10}$$

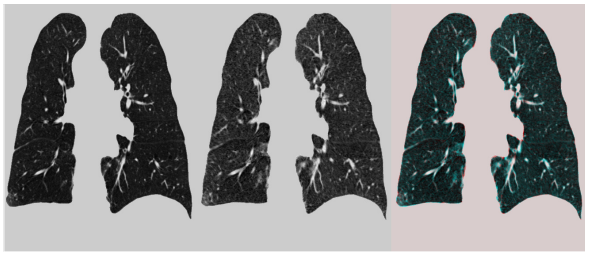
was calculated between the estimated geodesic distances in $\mathcal{D}$ and the Euclidean pairwise distances of the learned coordinates $y^{(\cdot)}$. Pearson's correlation coefficient ($\rho^2_{\mathcal{D},y}$) was computed between $\mathcal{D}$ and the distances of $y^{(\cdot)}$. The aim is to find the set of parameters d



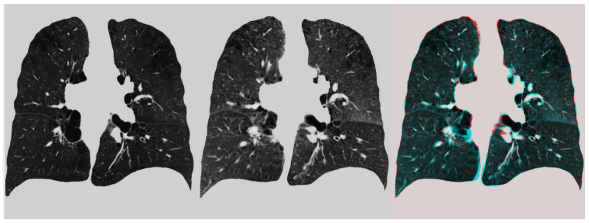
(a) Score 1



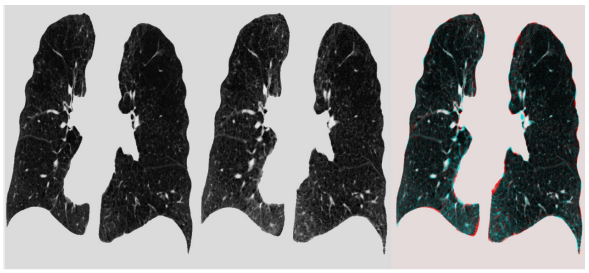
(b) Score 2



(c) Score 3



(d) Score 4



(e) Score 5

Figure 5.15: Illustrative registration results for the qualitative scoring system

and K, which best satisfy the relationship in Equation 5.7. This was performed by finding the pair of parameters K^* and d^* that minimises Equation 5.10 such that:

$$\arg\min_{K,d} \rho_{res} \tag{5.11}$$

For each embedding step (y^e, y^f and y^J), the combination of K and d that yielded the minimal residual variance for each embedding was determined with Equation 5.11. Grid-search parameters were set $d^* \in [1,5]$ and $K^* \in [5,100]$. This led to the parameters $K = [50,30,45]$ and $d = [5,5,4]$ for y^e, y^f and y^J respectively.

Two separate models were considered: 1) a model of the disease distributions (y^e, $y^f \rightarrow y^{c_1}$) and 2) a model of the disease and deformation distributions (y^e, y^f, $y^J \rightarrow y^{c_2}$). The first model (y^{c_1}) aimed to solely study the variation of emphysema and fSAD distributions in the COPD cohort. The second model (y^{c_2}) was constructed to investigate a more complete representation of the disease process by combining both emphysema and fSAD distributions with the local deformation distribution. Parameters for both models after application of Equation 5.11 were $K_{c_1} = 55$ and $K_{c_2} = 60$ with $d_{c_1} = 4$ and $d_{c_2} = 4$.

5.3.2.5 Single manifolds of emphysema, fSAD and deformation

Correlations between the individual embeddings and moments of the distributions were tested. This was performed by calculating Pearson's correlation coefficient between the first 3 dimensions of y^e, y^f and y^J with the following distribution parameters: median (φ), median absolute deviation (ρ), skewness (γ_1) and kurtosis (γ_2). The median and median absolute deviation were chosen over the mean and standard deviation due to the shape of the various disease and deformation distributions. After inspection, I concluded that there were 4 potential disease distribution shapes (normal, skew-normal, exponential, a mixture of normal distributions) and 3 potential deformation distributions (normal, skew-normal and a mixture of normal distributions). The median and median absolute deviation therefore better quantified the range of distributions.

The first and second components of the individual embeddings had strong to moderate correlations with the distribution parameters (Table 5.2). The first component had the strongest correlations whilst the second and third components of the embeddings were weaker but still mainly significant. For example, the first component of y^e correlates very strongly with all distribution parameters with weak, yet significant relationships with the second. For y^f, the first component strongly correlates with the median yet there is an equal

Table 5.2: Pearson correlation coefficient between embedding coordinates and the distributions using the median (φ), median absolute deviation (ρ), skewness (γ_1), kurtosis (γ_2). [$* = p < 0.05$, $\dagger = p < 10^{-3}$]. © 2017 Springer Nature. Reprinted, with permission from F. Bragman et al., Manifold learning of COPD, Medical Image Computing and Computer Assisted Interventions, 2017.

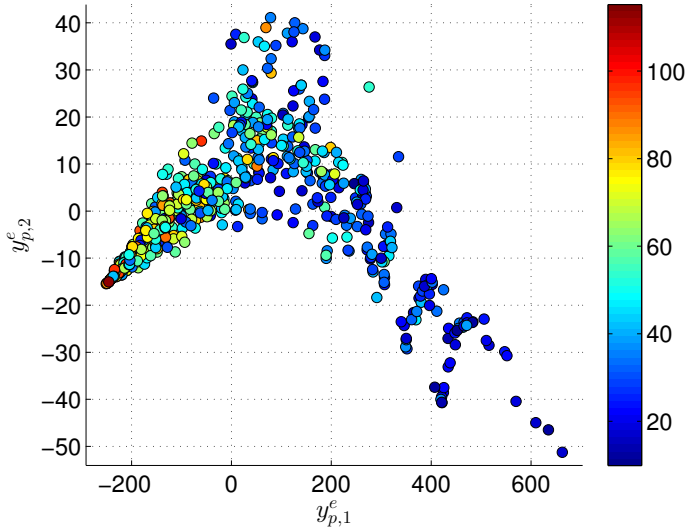
	PRM_{emph}			PRM_{fSAD}			J		
	y_1^e	y_2^e	y_3^e	y_1^f	y_2^f	y_3^f	y_1^J	y_2^J	y_3^J
φ	$0.96^\dagger$	$-0.19^\dagger$	0.01	$0.97^\dagger$	0.07	-0.01	$-0.48^\dagger$	-0.06	0.04
ρ	$0.89^\dagger$	$0.22^\dagger$	-0.00	$0.35^\dagger$	$-0.36^\dagger$	$-0.41^\dagger$	$-0.46^\dagger$	0.14^*	-0.09
γ_1	$-0.71^\dagger$	$-0.28^\dagger$	0.00	$-0.86^\dagger$	$0.21^\dagger$	$0.16^\dagger$	$-0.68^\dagger$	$-0.24^\dagger$	0.00
γ_2	$-0.41^\dagger$	$-0.26^\dagger$	-0.01	$-0.37^\dagger$	$0.33^\dagger$	$0.26^\dagger$	$-0.36^\dagger$	$-0.18^\dagger$	-0.01

correlation across the three components when considering the median absolute deviation and the kurtosis. These relationships demonstrate that manifold learning of the distributions modelled the inter-patient variations in the studied population.

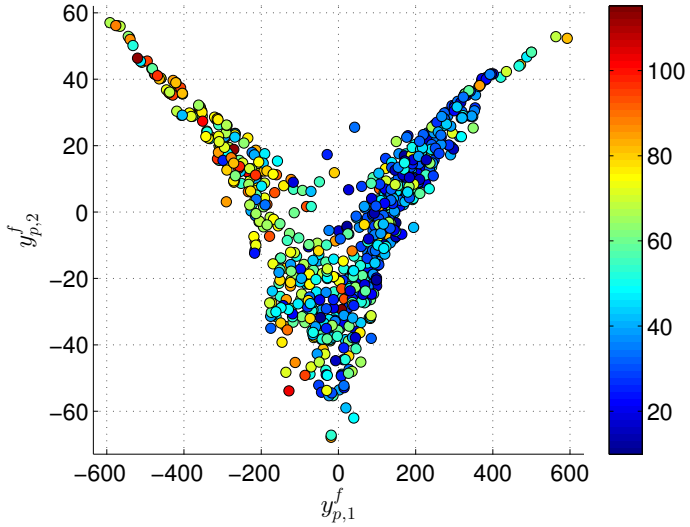
Two-dimensional projections of the underlying manifolds for emphysema, fSAD and the Jacobian distribution are visualised in Figure 5.16. They display the progression of the local-disease and deformation distributions correlated with FEV_1%predicted. A point of inflection is present in all manifolds, which correspond to significant shift in FEV_1%predicted. The lower correlation with the second component of the embeddings ($y_2^{(\cdot)}$) can be attributed to the variation of the distributions in this region. At lower levels of disease severity (FEV_1%predicted ≥ 60), the distributions mainly represent diffuse levels of disease with high homogeneity. As tissue destruction progresses and variations in lung deformation are more present, several modes in the distributions become apparent. As severity increases (FEV1 %predicted < 60), most distributions are seen to converge to similar states with small clusters of patients representing distinct patterns e.g. upper lobe dense emphysema. Detecting this point of inflection in the clinical management of COPD may be beneficial as the progression of disease may accelerate greatly past this point.

5.3.2.6 Association of models with disease severity

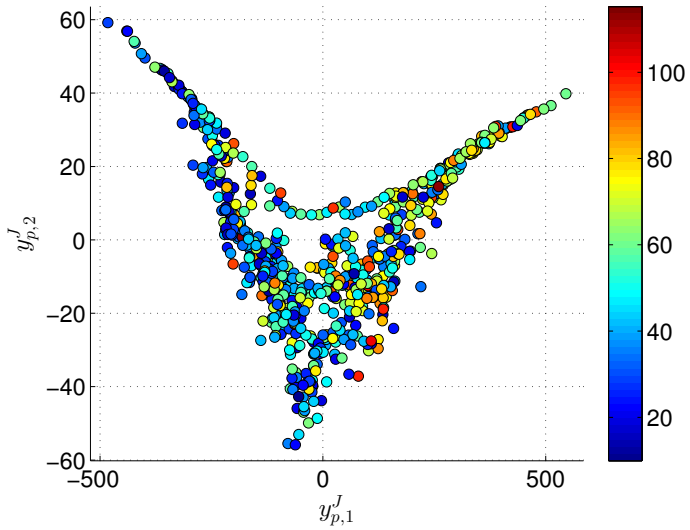
Two models were constructed to predict COPD severity using FEV_1%predicted and $FEV1/FVC$ (Table 5.3 and 5.4). The first model (y^{c_1}) is a fusion of the emphysema (y^e) and fSAD (y^f) manifolds whilst the second (y^{c_2}) also combines the deformation embedding (y^J). We considered three simple models (mean PRM_{emph}, mean PRM_{fSAD} and mean Jacobian $\mu(J)$) and compared them to univariate (Table 5.3) and multivariate (Table 5.4) models of



(a) Learned embedding of emphysema distributions - y^e



(b) Learned embedding of fSAD distributions - y^f



(c) Learned embedding of Jacobian distributions - y^J

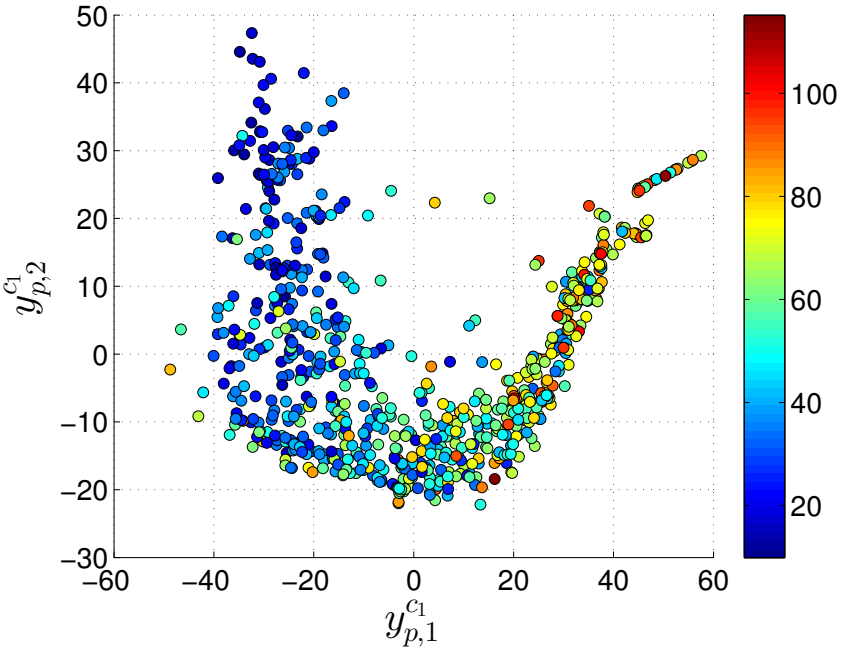
Figure 5.16: Embeddings of emphysema, fSAD and Jacobian distributions with FEV_1%predicted overlaid

embedding coordinates (y). The univariate models only consider the first component of $y^{(\cdot)}$ such that $Y = \beta_0 + \beta_1 \cdot y_1^{(\cdot)}$. The multivariate models make use of all dimensions of $y^{(\cdot)}$ where $Y = \beta_0 + \sum_i \beta_i \cdot y_i^{(\cdot)}$. Model performance was tested with the adjusted-r^2 and the Bayesian Information Criterion (BIC). It is defined as $\text{BIC} = \ln(n)k - 2\ln(\widehat{L})$, where n is the number of observed data points, k is the number of parameters in the model and $\widehat{L}$ is the maximised value of the likelihood function. The BIC is a criterion for model selection amongst a finite set of models ran on the same data. The model with the lowest BIC is preferred.

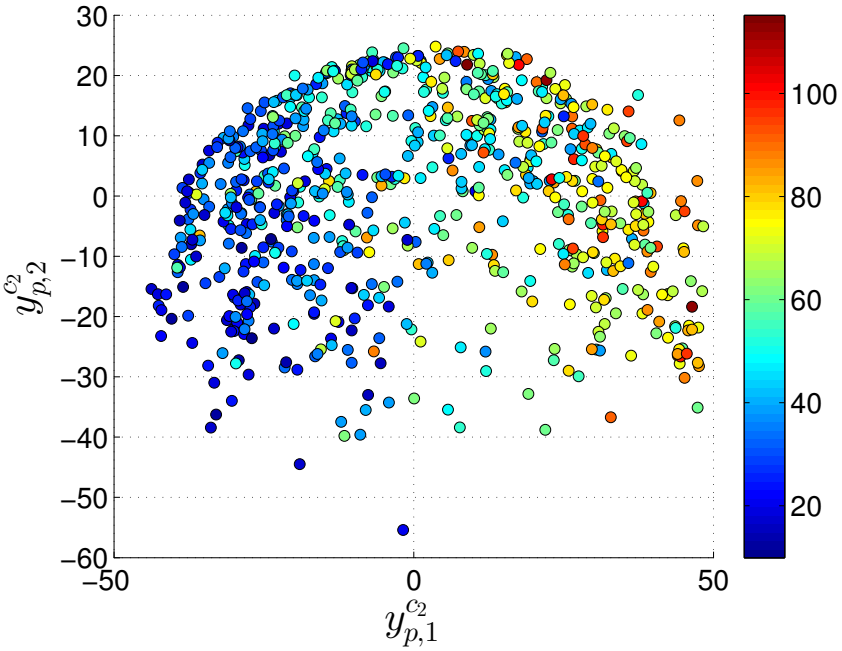
The univariate models ($y_1^{(e,f)}$) showed moderate improvement over the simple mean models. However, the combined models ($y_1^{c_1}$ and $y_1^{c_2}$) improved model prediction (Table 5.3). The multivariate model demonstrated best performance, with model 2 ($y^{c_2} = y^e + y^f + y^J$) performing best, even after adjusting for an increase in variables. It had a BIC of 620 compared to 625 (y^{c_1}) whilst the simple models had a BIC of 633, 650 and 648 for PRM_{emph}, PRM_{fSAD} and $\mu(J)$ respectively.

The increase in explanatory power was also seen when correlating the first component of the combined models ($y_1^{c_{1,2}}$) with FEV_1%predicted. The first components of the combined models had Pearson coefficients of $r = 0.67, p < 0.001$ and $r = 0.70, p < 0.001$ respectively. Coefficients for the mean models were $r = -0.63, p < 0.001$, $r = -0.50, p < 0.001$ and $r = 0.52, p < 0.001$.

To test whether manifold fusion based on mean differences of PRM_{emph}, PRM_{fSAD} and $\mu(J)$ performed better than manifold learning and fusion of the proposed distributions, I created a joint model between PRM_{emph} and mean PRM_{fSAD} in addition to a model with mean values of PRM_{emph}, PRM_{fSAD} with $\mu(J)$. In this case, the pairwise distances were mean differences between patients rather than the EMD between distributions. The same process described in Section 5.3.2.4 was applied to learn parameters K and d for the models. This lead to embeddings y^{mean_1} and y^{mean_2}. The correlation of the first component of each embedding was $r = 0.60, p < 0.001$ and $r = -0.65, p < 0.001$ in comparison to $r = 0.67, p < 0.001$ and $r = 0.70, p < 0.001$ for $y_1^{c_1}$ and $y_1^{c_2}$. This further corroborated the utility of manifold fusion of separate embeddings based on local disease and deformation distributions.



(a) Disease distribution embedding - y^{c_1}



(b) Disease and deformation distribution embedding - y^{c_2}

Figure 5.17: Combined models of emphysema, fSAD and Jacobian distributions with FEV_1%predicted overlaid. © 2017 Springer Nature. Reprinted, with permission from F. Bragman et al., Manifold learning of COPD, Medical Image Computing and Computer Assisted Interventions, 2017.

Table 5.3: Regression of models versus various clinical measures of COPD severity. Model performance quoted as adjusted-r^2. [$\dagger = p < 10^{-3}$]. © 2017 Springer Nature. Reprinted, with permission from F. Bragman et al., Manifold learning of COPD, Medical Image Computing and Computer Assisted Interventions, 2017.

	Mean features			$y_1^{(\cdot)}$				
Y	PRM_{emph}	PRM_{fSAD}	$\mu(J)$	$y_1^{c_1}$	$y_1^{c_2}$	y_1^{e}	y_1^{f}	y_1^{J}
FEV_1%predicted	$0.40^{\dagger}$	$0.25^{\dagger}$	$0.26^{\dagger}$	$0.45^{\dagger}$	$\mathbf{0.49}^{\dagger}$	$0.42^{\dagger}$	$0.29^{\dagger}$	$0.13^{\dagger}$
FEV_1/FVC	$0.51^{\dagger}$	$0.30^{\dagger}$	$0.22^{\dagger}$	$\mathbf{0.54}^{\dagger}$	$0.53^{\dagger}$	$0.54^{\dagger}$	$0.32^{\dagger}$	$0.09^{\dagger}$

Table 5.4: Regression of models versus various clinical measures of COPD severity. Model performance quoted as adjusted-r^2. [$\dagger = p < 10^{-3}$]. © 2017 Springer Nature. Reprinted, with permission from F. Bragman et al., Manifold learning of COPD, Medical Image Computing and Computer Assisted Interventions, 2017.

	$y^{(\cdot)}$				
Y	y^{c_1}	y^{c_2}	y^{e}	y^{f}	y^{J}
FEV_1%predicted	$0.48^{\dagger}$	$\mathbf{0.51}^{\dagger}$	$0.43^{\dagger}$	$0.34^{\dagger}$	$0.14^{\dagger}$
FEV_1/FVC	$0.59^{\dagger}$	$\mathbf{0.60}^{\dagger}$	$0.55^{\dagger}$	$0.38^{\dagger}$	$0.10^{\dagger}$

5.3.2.7 Trajectories of emphysema and fSAD progression

It is likely that trajectories of disease progression in COPD vary depending on the dominant disease phenotype. I assessed whether these can be modelled in the tissue disease model (y^{c_1}). I fitted trajectories of emphysema and fSAD progression in the manifold space of y^{c_1} to find distinct sets of patients where emphysema or fSAD is the dominant phenotype. To parameterise the space y^{c_1}, I used Nadayara-Watson kernel regression [145]. Kernel regression is a standard tool for determining a smooth, estimated function by using its noisy observations. A careful choice of the covariate (l) allows one to navigate the embedding space in various ways to test associations with many clinical metrics. For instance, if we were interested in studying the relationship between age of a patient and position in the learned manifold, patient age would be used as a covariate l in the regression. This would consequently provide a mechanism to study how the complex imaging features used in the manifold learning correlate with age.

The kernel regression is found by the following equation:

$$y^{c_1}(l(\cdot)) = \frac{1}{v}\sum_p K_G\left(\frac{l(\cdot)_p - l(\cdot)}{h}\right)\cdot y_p^{c_1} \tag{5.12}$$

where h was set to 1 and is a smoothing parameter called the bandwidth whilst K_G is a

Gaussian kernel of the form

$$K_G(x) = \frac{1}{\sqrt{(2\pi)}} \cdot \exp\left(-\frac{x^2}{2}\right). \tag{5.13}$$

The parameter v is a normalisation constant and is determined by

$$v = \sum_{p}^{P} K_G. \tag{5.14}$$

In the example of patient age regression, l is set to a vector $l \in [\text{age}_{min}, \text{age}_{min}+1, \cdots, \text{age}_{max}-1, \text{age}_{max}]$ where age_{min} and age_{max} are derived from the population. Equation 5.12 is then applied in the space of y^{c_1} for each increment of l. The Gaussian kernel in Equation 5.13 is used to calculate the average position within y^{c_1} where the bandwidth h controls the smoothness of the resulting curve.

In order to use the emphysema and fSAD disease distributions as covariates, I considered the Earth Movers Distance ($\mathcal{L}_{EMD}$) between the distributions of each patient and an idealised healthy distribution (Figure 5.18). The mean level of emphysema or fSAD within a patient may have seemed like a natural choice. However, this chapter has been arguing for more precise methodologies for measuring disease spread within the lung. To extract a scalar quantity from the proposed disease distributions, the *EMD* between a disease distribution and an idealised healthy distribution (H_{ideal}, Figure 5.18) was proposed. This computes how far away the distribution is from a purely healthy state and is analoguous to the mean level of emphysema or fSAD but for the proposed distributions. Two trajectories ($y^{c_1}(l(emph))$ and $y^{c_1}(l(fSAD))$) were quantified in the manifold space of y^{c_1} by application of Equation 5.12. These can be visualised Figure 5.19a.

I performed a naive classification of the patients in the space of y^{c_1} by considering the computed trajectories ($y^{c_1}(l(emph))$ and $y^{c_1}(l(fSAD))$). I hypothesised that a patient (p) is seen to follow an emphysema progression trajectory if the Euclidean distance from $y_p^{c_1}$ to $y^c(l(emph))$ is smaller than the distance to to $y^c(l(fSAD))$. At baseline, patients are classified as both emphysema and fSAD subtypes since progression of both distributions is in its infancy. This is visible as the black scatter points in Figure 5.19b. When considering the two sets of patients stratified by trajectory, the explanatory power of the embeddings improved in comparison to y^{c_1} (Table 5.4). The emphysema regression produced an adjusted-r^2 of 0.52 and 0.63 when predicting FEV_1%predicted and FEV_1/FVC respectively whilst fSAD was

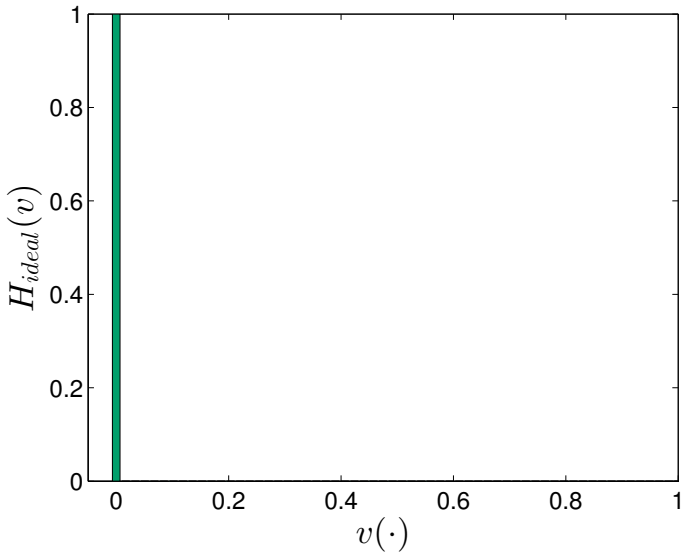


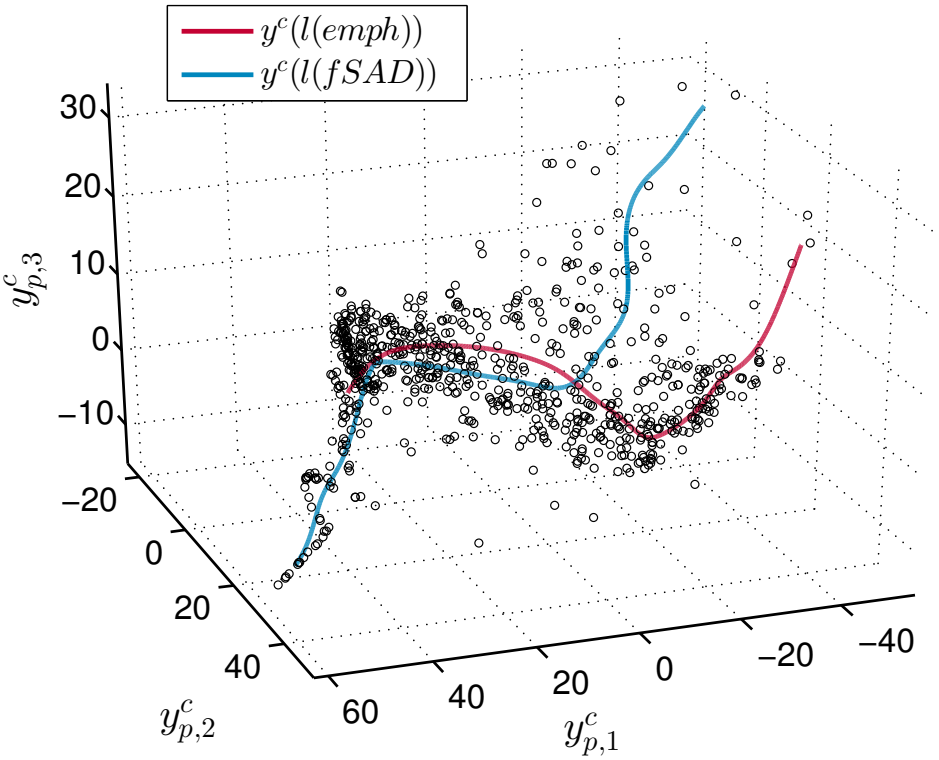
Figure 5.18: Theoretical healthy distribution for emphysema and fSAD.

0.45 and 0.62. In comparison, the adjusted-r^2 for y^{c_1} without patient stratification was 0.48 amd 0.59 (Table 5.4). This suggests that it is necessary to stratify patients first based on their expected trajectory of disease progression to better quantify response to therapy and COPD management.

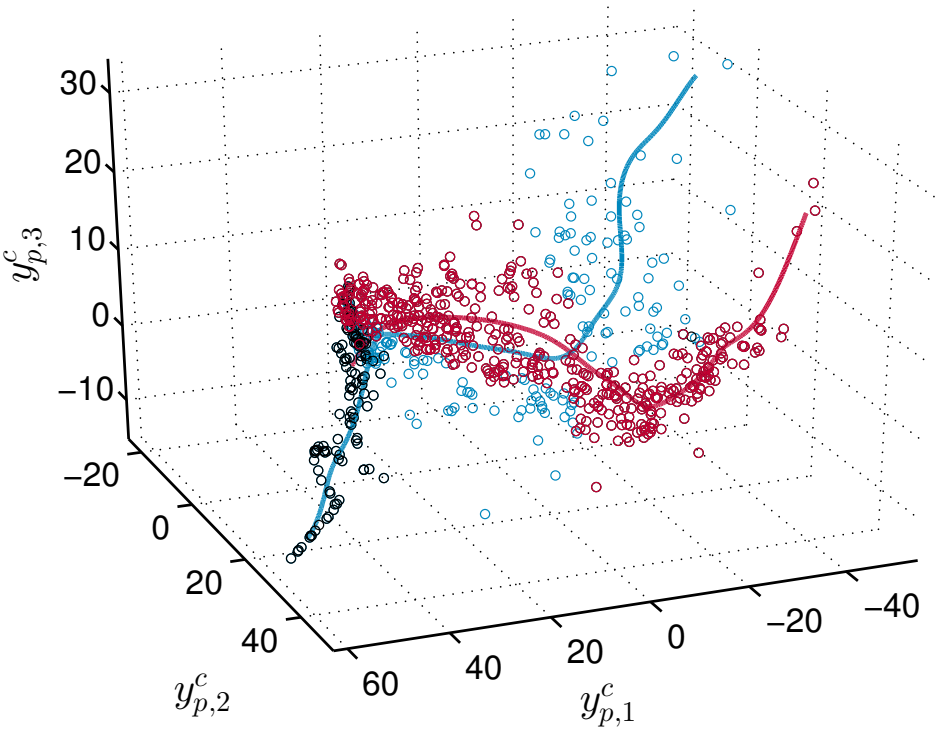
5.4 Discussion and conclusion

I have presented a method to model the spread of local features implicated in COPD progression. The disease distributions model local aspects of tissue destruction whilst parameterising global properties of heterogeneity and homogeneity. The deformation distribution quantifies the local effect of disease on lung function. Patients exhibiting different mechanisms of tissue destruction can have identical global averages yet can display different disease distributions. These variations are likely to cause differences in local biomechanical properties, which are captured by the deformation distribution.

I have shown that models of disease and deformation distributions are better predictors of COPD severity than conventional metrics. I have shown that embedding coordinates based on distribution dissimilarities have stronger correlations with $FEV_1\%$predicted than those learned from mean differences. Both these results suggest that the position of a patient in the manifold space of y^{c_1} or y^{c_2} is critical to the assessment of COPD progression. This is seen in the trajectory classification performed in y^{c_1}. Determining whether a patient is following a specific trajectory in the manifold space may help inform therapeutic decisions



(a) Disease distribution embedding - $y^{c_1} \in d = 3$



(b) Trajectory classification - $y^{c_1} \in d = 3$

Figure 5.19: Trajectories of emphysema and fSAD progression in y^{c_1}. © 2017 Springer Nature. Reprinted, with permission from F. Bragman et al., Manifold learning of COPD, Medical Image Computing and Computer Assisted Interventions, 2017.

and improve our understanding of the various COPD phenotypes.

There are various limitations associated with the proposed framework, which can be tackled to better model the complexities of COPD. The creation of the disease and deformation distributions was performed by sampling local ROIs across the lung. This disregarded potentially important lobar information. As discussed in Section 5.1, various studies have been performed to study whether upper-lobe and lower-lobe predominant disease have discordant pathophysiology. The framework applied in this Chapter could be extended with the lobe segmentation algorithm I presented in Chapter 4. It may be important to include lobar information in the processing. The deformation distribution for instance is likely to be influenced by its position within the lung since the lower lobes experience a greater level of deformation. The distributions can therefore be quantified on a lobar basis. Various models can then be investigated by varying the strategy for manifold fusion. The proposed distributions can also be extended to joint distributions by including gravitational height based on the dorsoventral axis. This circumvents the need for lobe segmentations. Adding regional information in the modelling either trough lobar models or joint distributions is likely to further increase the explanatory power of the models.

The nature of the ROI sampling can also be improved. Sampling the lung volume with overlapping ROIs can introduce a bias in the local measurements. At the interface of a large emphysematous cluster, the local measurement will not accurately reflect the underlying macrostructure since the local emphysema average will be 50%. There are various strategies that can help deal with this aspect. A random, sparse, selection of non-overlapping ROIs may help alleviate the bias. A better option may be to perform a supervoxel segmentation of the lung [1], as performed as a pre-processing step by Batmanghelich et al. [14]. This would subdivide the lung into coherent, spatially contiguous regions that could be used to constrain the sampling. Furthermore, classifying the image volume into emphysema and fSAD classes may potentially ignore important textural measures that can be used to detect specific emphysema subtypes [21, 245]. The local disease distributions provide a simple, intuitive way of modelling lung texture. However, critical information embedded within textural measures may add an important dimension to the analysis.

The method presented in this chapter has suggested that progression patterns of COPD may be extracted from cross-sectional data by the kernel regression in the manifold space of y^{c_1}. I have shown that lower-dimensional embeddings based on modelled inter-patient

distribution distances have stronger associations than traditional CT metrics with routinely used spirometric measures. More testing is necessitated to understand whether the trajectories represent distinct progression patterns. The longitudinal data, to be released by the COPDGene study [167] may be important in this experiment. Further experiments with various other clinical metrics such as the 6-minute walking distance, the BODE index (Body-mass, airflow Obstruction, Dyspnoea and Exercise), SGRQ (St-Georges Respiratory Questionnaire) and the DLCO (Diffusion capacity of the Lung for Carbon Monoxide) will present further evidence in the need to quantify the proposed distributions and go beyond traditional means of disease extent.

Chapter 6

Regional analysis of lung disease to study exacerbation susceptible COPD

∗ **Contributions**

I conceived the method, undertook and designed all the experiments and performed the data analysis. John Hurst provided support in the experiment design, feature selection and the presentation of results. Jamie McClelland and David Hawkes provided assistance in the presentation of the results, the data analysis and interpretation of the results.

Abstract

One of the principle aims of this thesis has been to develop tools that enable a more precise quantification of COPD from CT. Disease can spread regionally whilst the nature of local pathology may differ. This may affect biomechanical properties of the lung that contribute to airflow limitation. In order to capture these properties, a lobe segmentation algorithm was developed in Chapter 4 that enables me to perform lobar analysis of COPD whilst I presented the concept of local disease distributions in Chapter 5 to better model disease spread. These methods were combined to develop a pipeline of analysis that enables the quantification of different aspects of COPD at various anatomical scales.

The pipeline is presented within this chapter in addition to its application to the study of exacerbation susceptible COPD. Airway wall thickness has been shown to associate with exacerbation frequency yet there are conflicting reports on the relationship with emphysema. Associations with small airways disease has not been studied as of yet. It is not known whether regional differences differentiate susceptible and non-susceptible patients and whether measures of lung deformation vary with exacerbation frequency. The developed pipeline was applied to study these unknowns and determine whether regional analysis of lung CT improves the prediction of exacerbation frequency. I show within this chapter that regional analysis of lung CT does not help predict exacerbation frequency. However, I suggest new features that may improve associations between CT and spirometry for future imaging studies.

6.1 Introduction

Exacerbations of COPD are acute events that can dramatically alter the normal time-course of COPD [186]. They are a major determinant in the mortality, morbidity and quality of life of COPD patients [234] and cost the UK National Health Service over £253 million a year [26]. Reducing the frequency of exacerbations and improving the efficacy of therapeutic interventions related to exacerbations are key future targets in the management of COPD.

Susceptibility to exacerbations may be a distinct subtype of COPD. Although exacerbation frequency is seen to increase with COPD severity, there is an independent susceptibility phenotype [101] associated with higher severity of disease and a prior history of exacerbations. Patients can either be susceptible or non-susceptible to exacerbations of COPD. The current gold standard for diagnosis is based on a previous history of exacerbations. Currently, a threshold of two or more exacerbations per year is used to stage patients as exacerbation susceptible.

Computed Tomography (CT) has been increasingly used to develop quantitative tools [189] and study COPD [121] for precise phenotypic characterisation. Current use of quantitative CT applied to exacerbations has shown that higher levels of segmental airway wall thickness increase exacerbation frequency [87] whilst pulmonary arterial enlargement may be a strong predictor of exacerbations that cause hospitalisation [237]. The association between emphysema and exacerbations remains unclear. It has been reported that the occurrence of exacerbations accelerates emphysema progression in a two-year longitudinal study [205]. In contrast, results from the three-year ECLIPSE study suggested there is no relationship between exacerbations and emphysema progression [44]. This was observed in an independent study finding no correlation between emphysema extent and exacerbation frequency [69]. Independent of this, a recent study has shown that the use of convolutional neural networks can predict acute exacerbations from CT [76]. Further conflicting results were also present when assessing relationships between emphysema, airway wall thickening and exacerbation frequency [87]. A negative relationship between emphysema and exacerbation frequency was reported ($p < 0.001$) when the extent of emphysema was between 10% and 35%. In contrast, there was a marginally significant ($p = 0.047$) positive relationship between emphysema and exacerbations when the level of emphysema exceeded 35% [87].

It is now possible to quantify regional measurements and assess the extent of functional

small airways disease (fSAD).Functional small airways disease can now be measured using Parametric Response Mapping [70]. It is a distinct phenotype of COPD [96] and can predict lung function decline in mild-to-moderate stages of COPD [19]. Introducing fSAD as a component in CT-studies of COPD is likely to yield novel insights about mechanisms of disease.

Segmentation of the lobes facilitates regional analysis of disease extent. This had led to reports suggesting that the distribution of emphysema can vary in-terms of inter and intra-lobar heterogeneity with varying effects on pulmonary function [104, 215]. It is not known however if regional heterogeneity in emphysema and fSAD relate to exacerbations. Regional variations in disease extent may affect lung deformation, which can be captured using non-rigid registration of paired breath-hold scans. Regional analysis of lung deformation may reveal potential asymmetries in the lung function of exacerbation susceptible patients. Understanding potential regional differences may change the logic for lung volume reduction surgery (LVRS) using endobronchial valves [187]. It may also have potential repercussions for clinical end-points in strategies to manage exacerbations [173].

Within this chapter, I present an analysis pipeline that is applied to study exacerbation susceptibility but may also be applied to study other hypotheses related to COPD. The goal is to study the extent of disease and lung deformation at various anatomical scales. This is similar to the work of Murphy et al. [143] who analysed global and lobar features to study associations with spirometry. Disease is measured using the Parametric Response Mapping (PRM) technique, which allows classification of emphysema (PRM_{emph}) and fSAD (PRM_{fSAD}). Lung deformation is analysed through the Jacobian determinant, which measures fractional volume change on a voxel-wise basis. Moreover, the local disease distributions presented in Chapter 5 are quantified, which allows me to investigate the extent of diffuse and dense disease. These are introduced and defined in Section 6.2.1.4. By combining these measures from a global to a lobar level, this enables me to investigate whether regional measurements of lung disease and deformation help to better phenotype susceptible patients from CT.

6.2 Methods

There are various processing steps in the analysis pipeline. The lungs first need to be segmented (Section 6.2.1.1.1). The segmentations are then used to segment the lobes (Section 6.2.1.1.2). The segmentations are also used to help register scans acquired at inspiration to those at expiration (Section 6.2.1.2). Using results from the registration, a variety of features can be calculated. The Jacobian determinant can be obtained from the registration (Section 6.2.1.2) to analyse local volume change. The Parametric Response Mapping technique [71] can also be used to classify tissue as PRM_{emph} and PRM_{fSAD} by analysing the joint histogram of voxel intensities after registration (Section 6.2.1.3). The PRM classification can be exploited to create local disease distributions and obtain measures of diffuse and dense disease (Section 6.2.1.4). These features can be analysed at a global level using the lung segmentation or at a lobar level using the lobe segmentations.

6.2.1 Quantitative analysis

6.2.1.1 Lung and lobe segmentation

6.2.1.1.1 Lung segmentation

Lung masks were obtained using the Pulmonary Toolkit [52]. The Pulmonary Toolkit employs the algorithm of Hu et al. [100] to segment the lungs.

6.2.1.1.2 Lobe segmentation

The lobe segmentation algorithm described in Chapter 4 was employed. To deal with the volume of patients in the COPDGene dataset, I created a groupwise fissure prior from a subset of 300 patients using the same method as described in Chapter 4. To propagate the groupwise fissure prior to unseen patients in the COPDGene cohort, I registered each new patient to the groupwise space. I exploited the backwards transformation to resample the groupwise fissure into the unseen patient space and perform the lobe segmentation (Figure 6.1).

The lobe segmentation output is five separate lobar regions for analysis. This allowed me to compute upper-lobe, lower-lobe and inter-lobar metrics. Upper-lobe features were quantified using the weighted mean of the upper right and upper left lobe with the lobar volumes as weights. The same was applied for the lower lobes. Inter-lobar metrics were obtained by calculating the standard deviation of a feature across all five lobes. Since the lobes

are independent anatomical units, this will provide a measure of inter-lobar heterogeneity.

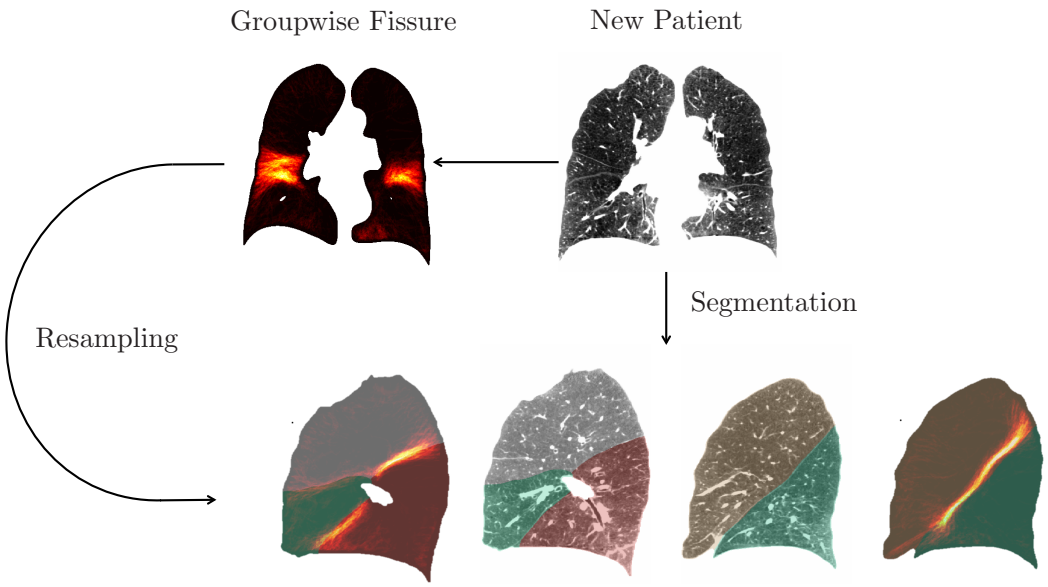


Figure 6.1: Lobe segmentation of a patient not included in the groupwise fissure prior. A new patient is registered to the groupwise space. The groupwise fissure prior is then resampled into the patient space to help guide the lobe segmentation in regions of incomplete fissures.

6.2.1.2 Nonrigid registration and deformation features

I registered scans acquired at at forced residual capacity (FRC) to those at total lung capacity (TLC). I used the diffeomorphic extension of the NiftyReg open-source registration platform [139]. The pipeline for registering paired breath-hold scans is the same as the one applied in Chapter 5.

The output of the registration is a transformation that maps every voxel in the TLC scan to the FRC scan. This transformation can be analysed to extract features that explain lung deformation. I computed the Jacobian determinant (J) from the deformation on a voxel-wise basis. It provides information on local volume change as the lung undergoes deformation from inspiration to expiration. It has a value ranging from zero to infinity. A Jacobian equal to one signifies no volume change. If the value is less than one, there is compression. If it is greater than one, there is expansion. I calculated the inverse of J at every voxel to analyse inspiration in the TLC reference frame. This was possible since the registration is diffeomorphic and symmetric.

I computed the mean and the standard deviation of J. The mean Jacobian ($\mu(J)$) quantifies the average level of volume change in the lung. The standard deviation of the Jacobian ($\sigma(J)$) quantifies heterogeneity of volume change. When $\sigma(J) = 0$, there is a

completely homogeneous pattern of volume change with no variation. The full range of analysed features are summarised in Table 6.1.

Table 6.1: Global and regional deformation features.

Region	Feature	Quantification
Global	$\mu(J)$ $\sigma(J)$	Average in lung mask
Upper-lobe	$\mu(J)$ $\sigma(J)$	Weighted mean of upper-lobe averages
Lower-lobe	$\mu(J)$ $\sigma(J)$	Weighted mean of lower-lobe averages
Inter-lobe	$\mu(J)$ $\sigma(J)$	Standard deviation of lobar averages

6.2.1.3 Tissue classification with Parametric Response Mapping

Voxel-wise classification of the lung was performed using Parametric Response Mapping [71]. A voxel was classified as emphysematous (PRM_{emph}) if it has an intensity below -950 Hounsfield Units (HU) at TLC and below -856 HU at FRC. A voxel was classified as functional small airways disease (PRM_{fSAD}) if it has an intensity at FRC below -856HU but greater than -950HU at TLC. The analysed PRM features are summarised in Table 6.2.

Table 6.2: Global and regional PRM features.

Region	Feature	Quantification
Global	PRM_{emph} PRM_{fSAD}	Average in lung mask
Upper-lobe	PRM_{emph} PRM_{fSAD}	Weighted mean of upper-lobe averages
Lower-lobe	PRM_{emph} PRM_{fSAD}	Weighted mean of lower-lobe averages
Inter-lobe	PRM_{emph} PRM_{fSAD}	Standard deviation of lobar averages

6.2.1.4 Quantifying the extent of diffuse and dense disease

The spread of emphysema in the lung and the appearance of small airways disease may vary from patient to patient. The progression of emphysema may display large regions of emphysematous bullae, emphysema progression that is diffuse or a mixture of both. I used the local disease distributions (Figure 6.2), presented in Chapter 5 to capture these

effects. These distributions go beyond global extents of disease (Figure 6.3). Extracting features from the disease distributions consequently allowed me to model more precisely the emphysema and fSAD spread in relation to exacerbation frequency.

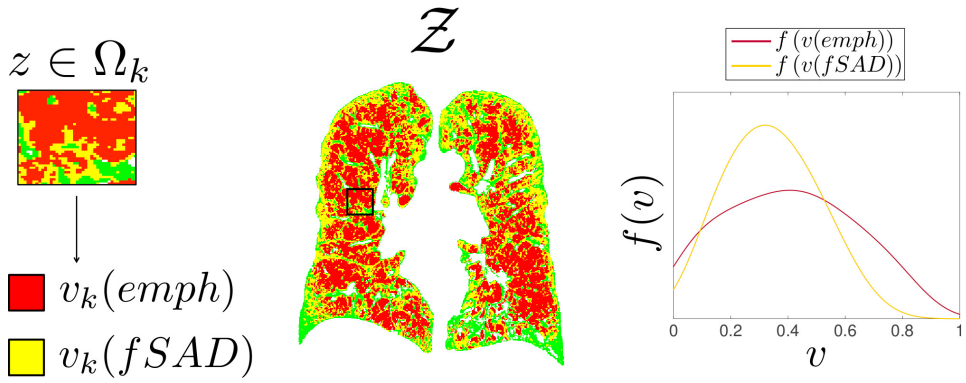


Figure 6.2: Quantification of local disease distributions. The aim is to extract more sensitive measures of emphysema and fSAD progression in the lung. The average amount of emphysema/fSAD provides context on the gross extent of disease. However, this measure does not describe whether the lung is composed mainly of regions of diffuse or dense disease. The local disease distribution aims to extract this information. 1) The lung classification volume output ($\mathcal{Z}$) from Parametric Response Mapping is sampled in local neighbourhoods to quantify the extent of emphysema/fSAD in small patches. 2) By sampling the lung volume in enough regions, a distribution of the measurements can be built. This distribution provides information on the extent of diffuse and dense disease.

I defined three states of tissue disease that enabled me to quantify more precisely the spread of PRM_{emph} and PRM_{fSAD} in the lung. Lung disease, measured locally in the lung ($v_k(\cdot)$) can be classified into three states (Figure 6.4):

$$v_k(\cdot) = \begin{cases} \text{diffuse} & \text{if } v_k < 0.30 \\ \text{intermediate} & \text{if } 0.30 \leq v_k \leq 0.60 \\ \text{dense} & \text{otherwise} \end{cases} \tag{6.1}$$

These thresholds were determined after qualitative analysis of local ROIs. The fraction of ROIs classified as diffuse, intermediate and dense within the lung provides complimentary information to the average extent of disease (Figure 6.5). The extent of diffuse disease is the percentage of diffuse classified regions and is defined as P_1. The same process is applied to intermediate disease to yield P_2 and dense disease for P_3 (Figure 6.5). The distribution features used are summarised in Table 6.3.

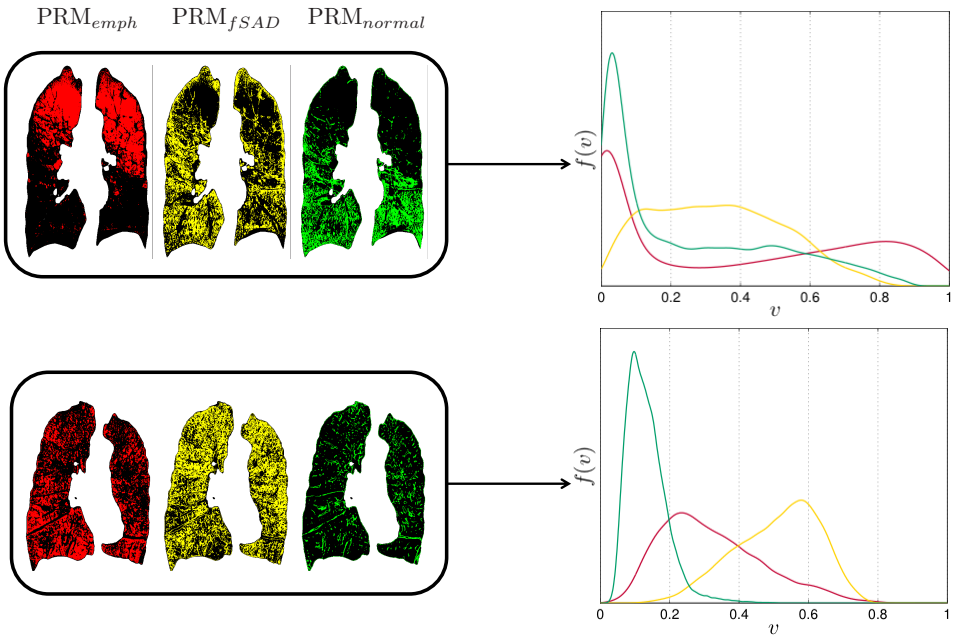


Figure 6.3: Illustration of two patients with an equal levels of emphysema in the lung (PRM$_{emph}$=25%) but with a different anatomical presentation. The lung in the top row has upper-lobe predominant emphysema, which is composed mainly of regions of dense disease ($v_k(emph) > 0.60$). In contrast, the bottom row displays a patient with a homogeneous presentation of emphysema. It mostly has regions of diffuse emphysema in the lung ($v_k(emph) \approx 0.20$) with limited regions of dense emphysema ($v_k(emph) > 0.60$).

Table 6.3: Global disease distribution parameters.

Region	Feature	Quantification
Global	$P_1(emph)$ $P_1(fSAD)$	Extent of diffuse disease in lung mask
	$P_2(emph)$ $P_2(fSAD)$	Extent of intermediate disease in lung mask
	$P_3(emph)$ $P_3(fSAD)$	Extent of dense disease in lung mask

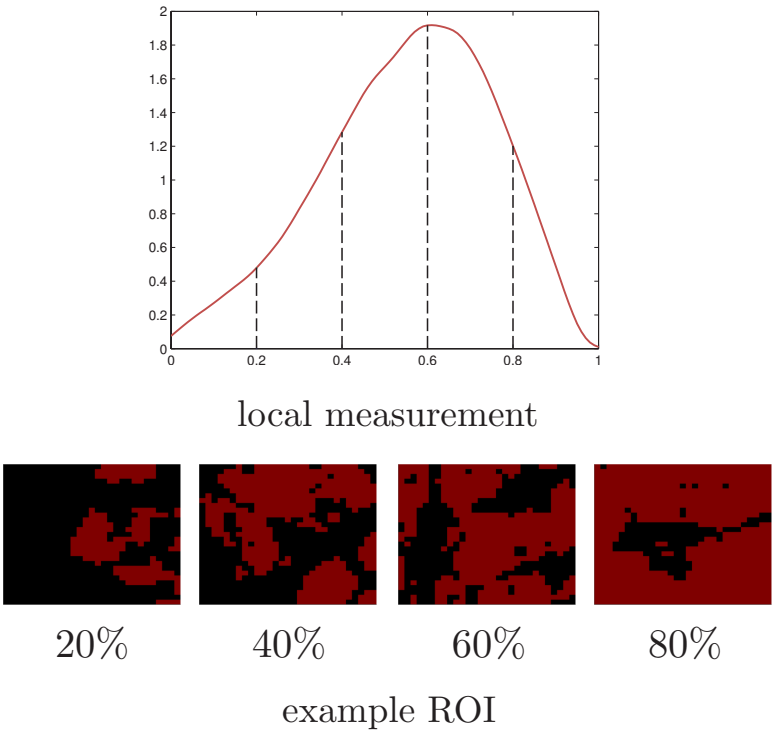
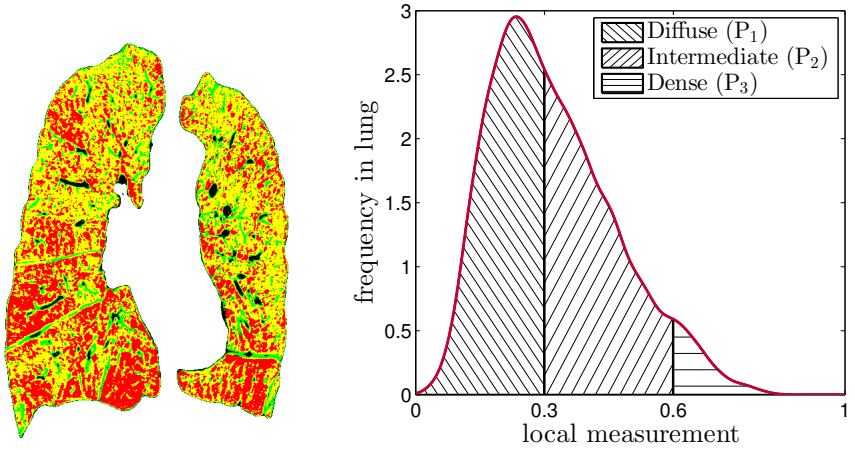


Figure 6.4: Illustration of PRM_{emph} or PRM_{fSAD} measured locally and its position in the local disease distribution. As the local measurement (x-axis) shifts from 0 to 1, the density of the measured feature in that location increases. Given a local measurement less than 0.3 (e.g. 30% of PRMemph in the local neighbourhood), I define the disease as diffuse. From 0.30 to 0.60, the disease is a mixture of both diffuse and dense and is termed intermediate. At any regions with $>$ 0.60, the disease is dense. The y-axis represents the probability i.e. frequency of region within the lung.



(a) PRM classification

(b) Local emphysema distribution of Figure 6.5a: $P_1 = 0.51$, $P_2 = 0.44$ and $P_3 = 0.05$

Figure 6.5: Illustration of diffuse (P_1), intermediate (P_2) and dense (P_3) disease features. To quantify the extent of diffuse, intermediate and dense disease in the lung, the percentage of regions classified as diffuse, intermediate and dense are calculated. 1) Percentage of diffuse disease (P1): % v <0.30. 2) Percentage of intermediate disease (P2): % regions $0.30 \leq v \leq 0.60$. 3) Percentage of dense disease (P3): % regions $v > 0.30$)

6.2.2 Statistical analysis

Statistical analysis was performed using MATLAB 2013b[1] and Python StatsModels statistical library[2]. Group statistics between continuous variables were quantified by computing the group mean and standard deviations. Significance between groups of continuous variables was assessed with a two-sample t-test with the assumption that the samples stem from normal distributions with unequal variances. For categorical variables, a chi-squared (χ^2) test of independence was computed.

A negative binomial regression with a log-link was used to model associations between the CT-features and exacerbation frequency. I accounted for scanner variability with dummy variables for each scanner type. I used post-bronchodilator FEV_1%predicted as a confounder. Since FEV_1%predicted is obtained by normalising the raw FEV_1 with respect to age, height, gender and ethnicity, no further confounding variables were used. The regression was also performed without FEV_1%predicted as a confounding variable. The regression coefficient for a feature when FEV_1%predicted was used was denoted as β_{FEV_1} and was β without in the table heading. The standard error for the coefficient was reported. Associations between the CT-features and FEV_1%predicted was performed by calculating Pearson's linear correlation coefficient (r).

The independent variables (X_i) in the regression were all standardised such that they had zero mean and unit standard deviations. This allowed for a direct comparison of the regression coefficients (β_i). A unit standard deviation increase in X_i leads to a fold-change of $\exp(\beta_i)$ in exacerbation frequency (Y). Since the variables all have unit standard deviations after transformation, the magnitude of β_i reflected the relative importance of the feature in the prediction of Y within a multivariate model.

6.3 Results

6.3.1 Study population

Subjects participating in COPDGene (Genetic Epidemiology of COPD), a large multi-centre observational cohort study were used in the analysis. Current and former smokers with greater than or equal to 10 pack-year smoking history, with and without airflow obstruction were enrolled.

I downloaded 1,154 COPD patient scans from the COPDGene database that had paired

[1] www.mathworks.com
[2] www.statsmodels.org

breath-hold scans and were acquired using the STANDARD reconstruction kernel. I manually inspected all the registration results and the lobe segmentation results. The manual quality control of the registrations was performed according to the qualitative scoring system presented in Chapter 5, Section 5.3.2.3.[3] Registrations were rejected when there were major errors close at the fissures or close to the lung diaphragm.

Quality control of the lobe segmentations was also qualitative. However, no scoring system was devised the measure errors. This is because segmentation failures occured when there were errors in the airway labelling that could not be corrected or when the algorithm failed at extrapolating the fissure boundaries in cases with significantly large areas of incomplete fissures. As a result, patients with errors were easily rejected. I rejected 94 subjects resulting in 675 patients used for the subsequent analysis. As a sanity check, the mean level of emphysema in the upper lobes and in the lower lobes in the reamining patients was calculated. Pearson's linear correlation coefficient between these measures and the data supplied by COPDGene (obtained through VIDA diagnostics) was calculated. A coefficient of 0.99, $p < 10^{-5}$ was obtained demonstrating correct lobar segmentations in the final dataset.

6.3.2 Subject demographics

The baseline characteristics of the exacerbation susceptible and non-susceptible patients are reported in Table 6.4. This showed as expected, that the susceptible group have much more severe lung function impairment, demonstrated by significantly smaller spirometric scores and a higher proportion of GOLD 3 and 4 patients. I also found that women are more likely to be exacerbation susceptible (58.1% versus 42.8%, $p = 0.001$). Within this cohort, the patients who were exacerbation susceptible were younger (mean age, 62.4 versus 64.3 years, p=0.017). I found no differences in smoking history (number of packs per year) between both groups (mean, 54.5 versus 54.8, p=0.89).

[3]The results from the registration quality control in Section 5.3.2.3 were obtained on the same dataset analysed within this chapter. Consequently, statistics from Section 5.3.2.3 are applicable to this study.

Table 6.4: Subject demographics

Parameter	Non-Susceptible	Susceptible	p value
Subjects, n	538	136	
Age, yr	64.3 (8.5)	62.4 (8.2)	0.017
Sex, (%)			
Male	57.2	41.9	0.001
Female	42.8	58.1	
GOLD classification, n (%)			
Stage 1	69 (12.8)	3 (2.2)	
Stage 2	220 (40.9)	41 (30.1)	$< 10^{-5}$
Stage 3	162 (30.1)	49 (36.0)	
Stage 4	87 (16.1)	43 (31.6)	
Smoking history, pack-years	54.5 (27.8)	54.8 (24.1)	0.89
Exacerbations, n/yr	0.27 (0.44)	2.97 (1.25)	-
FEV_1, L	1.54 (0.73)	1.15 (0.58)	$< 10^{-5}$
FEV_1% predicted	53.5 (21.9)	41.8 (18.5)	$< 10^{-5}$
FVC, L	3.02 (1.01)	2.62 (0.96)	$< 10^{-5}$
FVC % predicted	79.8 (20.2)	72.1 (19.9)	$< 10^{-5}$
FEV_1 / FVC	0.50 (0.13)	0.43 (0.13)	$< 10^{-5}$

6.3.3 From global to regional analysis of lung CT

In this section, I investigated associations between the proposed CT-features and exacerbation frequency. For each feature and at each scale of analysis, I performed a two-sample t-test between the non-susceptible and susceptible group. This result was reported as p-value in the results table. I then performed a regression between the proposed feature and exacerbation frequency with the resulting coefficient β. This was then repeated with FEV_1%predicted as a confounding variable to account for FEV_1 but also patient age, sex and height. This lead to the regression coefficient for the feature as β_{FEV_1}. The correlation between the feature and FEV_1%predicted was calculated with Pearson's linear correlation coefficient coefficient and reported as r.

6.3.3.1 Global measures of lung disease and deformation

The susceptible group of patients represents a cohort with more severe COPD. This was exemplified in the global extent of emphysema (16.4% versus 12.3% p=0.004, Table 6.5), the mean volume change (1.32 versus 1.37 $p =< 10^{-3}$, Table 6.6) and also heterogeneity

of volume change (0.26 versus 0.30 p=0.01,$\sigma(J)$, Table 6.6). These all showed significant associations with exacerbation frequency without accounting for FEV_1%predicted. This suggested that within this cohort, an exacerbation susceptible lung could be visualised as a lung with more emphysema with less volume change that is more homogeneous. However, with the addition of FEV_1%predicted, these relationships were lost since FEV_1%predicted strongly predicted frequency in this cohort but also because PRM_{emph}, PRM_{fSAD}, $\mu(J)$ and $\sigma(J)$ were moderately to strongly correlated with FEV_1%predicted suggesting multicolinearity effects.

Table 6.5: Global results for PRM_{emph} and PRM_{fSAD}. β_{FEV_1} is the regression with FEV_1%predicted as a confounder and is β without. The correlation between the parameter and FEV_1%predicted is r. [$* : p < 0.05$, $\dagger : p < 10^{-3}$]

Parameter	$n/yr < 2$	$n/yr \geq 2$	p-value	β	β_{FEV_1}	r
PRM_{emph}	12.3 (12.7)	16.4 (15.2)	0.004	0.20 (0.06)†	-0.11 (0.07)	$-0.64^\dagger$
$PRMf_{fSAD}$	29.1 (13.1)	30.8 (11.5)	0.120	0.12 (0.06)	-0.14 (0.08)	$-0.51^\dagger$

Table 6.6: Global results for $\mu(J)$ and $\sigma(J)$. β_{FEV_1} is the regression with FEV_1%predicted as a confounder and is β without. The correlation between the parameter and FEV_1%predicted is r. [$* : p < 0.05$, $\dagger : p < 10^{-3}$]

Parameter	$n/yr < 2$	$n/yr \geq 2$	p-value	β	β_{FEV_1}	r
$\mu(J)$	1.37 (0.17)	1.32 (0.13)	$< 10^{-3}$	-0.22 (0.07)†	0.08 (0.08)	$0.54^\dagger$
$\sigma(J)$	0.30 (0.13)	0.26 (0.14)	0.01	-0.19 (0.07)*	-0.01 (0.07)	$0.38^\dagger$

6.3.3.2 Global measure of local disease distribution

The creation of the local disease distributions (Section 6.2.1.4) and the resulting parameterisation allowed me to go beyond global measures of PRM_{emph} and PRM_{fSAD}. The proposed P_1, P_2 and P_3 parameters measured the extent of diffuse, intermediate and dense $PRM_{emph/fSAD}$.

The results in Table 6.7 and 6.8 demonstrated their assocations with exacerbation susceptibility, showling largely negative results when attempting to predict exacerbation susceptibility.

There were significant differences between the mean levels of the features across both populations such as in diffuse ($P_1(emph)$), intermediate ($P_2(emph)$) and dense emphysema ($P_3(emph)$) (Table 6.7). However, these differences were caused by the higher proportion

of severe COPD patients in the susceptible group.

Importantly, all features correlated significantly with FEV_1%predicted (r, Table 6.7 and 6.8), suggesting their potential to be used in future imaging studies. Additionally, $P_2(fSAD)$ and $P_3(fSAD)$ were associated with exacerbation frequency independent of FEV_1%predicted (Table 6.8) although the polarity of the regression coefficient β_{FEV_1} was reversed from β.

Table 6.7: Global results for $f(v(emph))$. β_{FEV_1} is the regression with FEV_1%predicted as a confounder and is β without. The correlation between the parameter and FEV_1%predicted is r. [$* : p < 0.05$, $\dagger : p < 10^{-3}$]

Parameter	$n/yr < 2$	$n/yr \geq 2$	p-value	β	β_{FEV_1}	r
$P_1(emph)$	0.85 (0.20)	0.79 (0.25)	0.009	-0.17 (0.06)*	0.11 (0.07)	0.60†
$P_2(emph)$	0.11 (0.14)	0.14 (0.16)	0.044	0.14 (0.06)*	-0.17 (0.07)	$-0.60^\dagger$
$P_3(emph)$	0.04 (0.07)	0.07 (0.11)	0.003	0.19 (0.06)†	-0.01 (0.06)	$-0.47^\dagger$

Table 6.8: Global results for $f(v(fSAD))$. β_{FEV_1} is the regression with FEV_1%predicted as a confounder and is β without. The correlation between the parameter and FEV_1%predicted is r. [$* : p < 0.05$, $\dagger : p < 10^{-3}$]

Parameter	$n/yr < 2$	$n/yr \geq 2$	p-value	β	β_{FEV_1}	r
$P_1(fSAD)$	0.54 (0.28)	0.51 (0.26)	0.10	-0.11 (0.06)	0.16 (0.07)*	0.53†
$P_2(fSAD)$	0.40 (0.23)	0.44 (0.21)	0.02	0.15 (0.06)*	-0.16 (0.08)*	$-0.58^\dagger$
$P_3(fSAD)$	0.06 (0.11)	0.06 (0.11)	0.50	-0.04 (0.06)	-0.09 (0.07)	$-0.14^\dagger$

6.3.3.3 Lobar measures of lung disease and deformation

This section presents analysis of the lobar features and their associations with exacerbation susceptibility and FEV_1%predicted. Analysis of the regional features showed various differences across groups, demonstrating the potential utility of lobar-based measurements of disease. Regression results showed that these features do not help predict exacerbation susceptibility when taking into account the degree of lung function impairment.

I found significant regional differences between the susceptible and non-susceptible group. Lower lobe PRM_{fSAD} was elevated in the susceptible group (0.28 versus 0.24 $p = 0.004$, Table 6.9) and was more strongly correlated with FEV_1%predicted than upper-lobe PRM_{fSAD} (r =-0.60, $p < 10^{-3}$ versus r =-0.40, $p < 10^{-3}$, Table 6.9). The inter-lobar metrics of PRM_{emph} and PRM_{fSAD} suggested a more heterogeneous presentation of emphy-

sema (0.07 versus 0.05 $p < 10^{-3}$, Table 6.11) yet more homogeneous fSAD across the lobes (0.07 versus 0.08 $p = 0.022$, Table 6.11).

As observed in the previous sections (Section 6.3.3.1 and 6.3.3.2), these differences were primarily due to differences in COPD severity across groups. The regression (β) demonstrated that most regional features do not associate with exacerbation frequency when adjusting for FEV_1%predicted. Conflicting results were observed when associations were tested without adjusting for lung function impairment. Upper-lobe features of lung deformation did not associate with frequency but lower-lobe features did (Table 6.10). Furthermore, there was an association with inter-lobar $\mu(J)$ (Table 6.12). This also supported the idea the susceptible lungs experience less expansion that is more homogeneous in the lower lobes.

When adding FEV_1%predicted to control for spirometric severity, only lower lobe PRM_{emph} (Table 6.9), upper lobe PRM_{fSAD} (Table 6.9) and upper-lobe $\mu(J)$ (Table 6.10) associated with exacerbation frequency. A switch in the polarity of regression coefficient from β to β_{FEV_1} was also observed.

Table 6.9: Upper and lower lobe results for PRM_{emph} and PRM_{fSAD}. β_{FEV_1} is the regression with FEV_1%predicted as a confounder and is β without. The correlation between the parameter and FEV_1%predicted is r. [$* : p < 0.05$, $\dagger : p < 10^{-3}$]

Parameter	Region	$n/yr < 2$	$n/yr \geq 2$	p-value	β	β_{FEV_1}	r
PRM_{emph}	Upper Lobe	0.14 (0.14)	0.19 (0.18)	$< 10^{-3}$	0.20 (0.06)*	-0.06 (0.07)	$-0.53^\dagger$
	Lower Lobe	0.10 (0.12)	0.14 (0.14)	0.005	0.17 (0.06)*	-0.15 (0.07)*	$-0.62^\dagger$
PRM_{fSAD}	Upper Lobe	0.32 (0.13)	0.32 (0.12)	0.81	0.01 (0.06)	-0.20 (0.07)†	$-0.40^\dagger$
	Lower Lobe	0.24 (0.14)	0.28 (0.13)	0.004	0.21 (0.06)†	-0.07 (0.08)	$-0.60^\dagger$

Table 6.10: Upper and lower lobe results for $\mu(J)$ and $\sigma(J)$. β_{FEV_1} is the regression with FEV_1%predicted as a confounder and is β without. The correlation between the parameter and FEV_1%predicted is r. [$* : p < 0.05$, $\dagger : p < 10^{-3}$]

Parameter	Region	$n/yr < 2$	$n/yr \geq 2$	p-value	β	β_{FEV_1}	r
$\mu(J)$	Upper Lobe	1.31 (0.16)	1.28 (0.14)	0.11	-0.08 (0.06)	0.16 (0.07)*	$0.43^\dagger$
	Lower Lobe	1.46 (0.23)	1.38 (0.18)	$< 10^{-3}$	-0.28 (0.07)†	-0.02 (0.08)	$0.55^\dagger$
$\sigma(J)$	Upper Lobe	0.22 (0.11)	0.20 (0.11)	0.05	-0.10 (0.06)	0.08 (0.07)	$0.35^\dagger$
	Lower Lobe	0.31 (0.14)	0.27 (0.15)	0.002	-0.26 (0.07)†	-0.08 (0.07)	$0.37^\dagger$

Table 6.11: Inter-lobar results for PRM_{emph} and PRM_{fSAD}. β_{FEV_1} is the regression with FEV_1%predicted as a confounder and is β without. The correlation between the parameter and FEV_1%predicted is r. [$* : p < 0.05$, $\dagger : p < 10^{-3}$]

Parameter	$n/yr < 2$	$n/yr \geq 2$	p-value	β	β_{FEV_1}	r
PRM_{emph}	0.05 (0.05)	0.07 (0.06)	$< 10^{-3}$	0.17 (0.06)*	-0.04 (0.07)	$-0.51^\dagger$
PRMf_{fSAD}	0.08 (0.04)	0.07 (0.03)	0.022	-0.20 (0.07)*	-0.08 (0.07)	$0.25^\dagger$

Table 6.12: Inter-lobar results for $\mu(J)$ and $\sigma(J)$. β_{FEV_1} is the regression with FEV_1%predicted as a confounder and is β without. The correlation between the parameter and FEV_1%predicted is r. [$* : p < 0.05$, $\dagger : p < 10^{-3}$]

Parameter	$n/yr < 2$	$n/yr \geq 2$	p-value	β	β_{FEV_1}	r
$\mu(J)$	0.14 (0.07)	0.13 (0.08)	0.08	-0.14 (0.07)*	0.01 (0.07)	$0.32^\dagger$
$\sigma(J)$	0.08 (0.05)	0.08 (0.08)	0.63	-0.08 (0.07)	-0.01 (0.06)	$0.14^\dagger$

6.3.4 Models of exacerbation frequency

6.3.4.1 Experiment design

In this section, I investigated whether the regional analysis of lung CT in addition to the proposed measures of disease distribution measures helped improve the prediction of exacerbation frequency. This was performed by iteratively increasing the complexity of the model by including measures of lung deformation, lobar features and the extent of diffuse and dense disease (Table 6.13).

Table 6.13: Nested models to evaluate associations with exacerbation frequency. Global measures are features that were calculated within the lung mask. Inter-lobar represents the standard deviation of lobar features and lobar represents upper and lower lobe values. $P_{1/2/3}$ represent the features of diffuse, intermediate and dense disease across the lung.

Model	Features					
	Global PRM	Global J	Inter-lobar PRM	Lobar PRM	Lobar J	$P_{1/2/3}$
1	✓					
2	✓	✓				
3	✓	✓	✓			
4	✓	✓		✓		
5	✓	✓	✓		✓	
6	✓	✓	✓		✓	✓

Model performance was assessed using Akaikes Information Criterion (*AIC*). The *AIC*

penalises the log-likelihood of the regression with the number of independent variables in the model. The *AIC* is defined as: $AIC = 2k - 2ln(\hat{L})$ where k is the number of variables and $\hat{L}$ is the maximised value of the likelihood function. A smaller AIC represents a better model.

As a baseline model of performance (Model 0), I considered a model with only $FEV_1\%$ predicted. I used dummy variables to account for scanner variability in all models. Since the AIC penalises the number of parameters, I also used dummy variables in Model 0.

6.3.4.2 Results

I observed that $FEV_1\%$ predicted was the strongest predictor of exacerbation frequency in all models (Table 6.14). I also observed that the addition of CT-based features did not significantly improve the prediction. This includes traditional features such as mean levels of emphysema but also the more complex metrics presented in this chapter. Despite iteratively adding more complex features, model performance did not increase although this may have been biased by naively adding non-significant features in the iterative construction of the nested models. The best model performance was seen in Model 4 (Table 6.13), which considered global measures of PRM_{emph}, PRM_{fSAD}, $\mu(J)$, $\sigma(J)$, and upper and lower lobar measures of PRM_{emph} and PRM_{fSAD} although none of these variables were significant. However, this is not significant since model 0 had an AIC equal to 1639.

Table 6.14: Performance of the nested models to evaluate associations with exacerbation frequency. Only significant features at $p < 0.05$ are reported. [$* : p < 0.05$, $\dagger : p < 10^{-5}$]

Model	*AIC*	Features - β (SE)
0	1639	$FEV_1\%$predicted†: -0.63 (0.09)
1	1637	$FEV_1\%$predicted†: -0.62 (0.10)
2	1640	$FEV_1\%$predicted†: -0.60 (0.10)
3	1642	$FEV_1\%$predicted†: -0.57 (0.10)
4	1636	$FEV_1\%$predicted†: -0.57 (0.10)
5	1641	$FEV_1\%$predicted†: -0.62 (0.10) Global $\sigma(J)^{*}$: -0.58 (0.28)
6	1643	$FEV_1\%$predicted†: -0.60 (0.10) Global $\sigma(J)^{*}$: -0.64 (0.30) Inter-lobar $PRM_{fSAD}{}^{*}$: -0.18 (0.09)

6.4 Discussion

Within this chapter, I presented a pipeline of analysis that aimed to extract features to model various aspect of lung disease and lung deformation. The framework can be used to analyse various anatomical scales to capture global and regional aspects of lung pathophysiology. The pipeline can be used to study various clinical hypotheses related to COPD and other progressive lung diseases. This includes finding better associations between FEV_1 and CT or studying the effect of lower lobe diffuse and dense emphysema on markers of COPD severity and mortality.

The analysis performed within this chapter was applied to exacerbation susceptible COPD. Understanding potential structural and functional differences between susceptible and non-susceptible patients is important. Whilst the current gold standard for future exacerbations is a past history of exacerbations, identifying features that are sensitive to structural changes in lung tissue and associate with exacerbation frequency may help identify patients for targeted research and therapeutic development.

Previous efforts in studying potential radiological phenotypes of COPD exacerbations have focused on the global extent of emphysema [87, 205, 44, 69] and airway-related measurements [87]. I have extended these studies by including Parametric Response Mapping (PRM), which includes emphysema and an indirect measure of functional small airways disease (fSAD) in addition to two measures of lung deformation. These features were analysed at the whole lung level and also at a lobar level (Tables 6.1, 6.2 and 6.3). This enabled me to investigate potential regional associations with exacerbation frequency that may not be present a global level. This is to my knowledge the first study that has analysed these features at various anatomical scales in relation to exacerbation susceptibility.

Throughout the analysis, I found significant differences in various features across the susceptible and non-susceptible groups. It was found that levels of diffuse emphysema were reduced across the susceptible group yet dense emphysema was increased (Table 6.7). Another example relates to the significant differences observed across groups when considering the biomechanics of the lung on a regional basis (Table 6.12). However, these results can be attributed to a class imbalance in the non-susceptible and susceptible cohorts. There was a higher proportion of GOLD 3 and 4 patients in the susceptible group ($p < 10^{-5}$) (Table 6.4). Since the majority of the studied imaging features correlated strongly with FEV_1%predicted, this biased the analysis. When accounting for FEV_1%predicted in the

regression of exacerbation susceptibility, no significant associations were found.

The experiment in Section 6.3.4 was performed to test whether the addition of regional features added significantly to the prediction of exacerbation frequency when controlling for global levels. This was performed by constructing nested models and iteratively including more complex features. This experiment also yielded negative results, corroborating the results seen when assessing the features individually. The main observation is that FEV_1%predicted most strongly predicts exacerbation frequency and that imaging features may not be useful. However, recent work by Gonzalez et al. [76] has shown that deep learning of chest CT may be predictive of exacerbations. Further work is needed to interpret their trained network to understand the limitations of the designed features proposed within this chapter.

Despite the lack of positive findings, use of the pipeline has demonstrated the need to go beyond global averages. Various features that quantified the nature of disease spread (diffuse versus dense emphysema) and the extent of lobar disease correlated strongly with spirometry and demonstrated significant differences across cohorts. At a global level, mean emphysema was elevated whilst volume change and deformation heterogeneity were reduced in the susceptible group. However, it was then seen that only lower-lobe measures of lung deformation in addition to lower-lobe fSAD varied across groups. The need to go beyond the global averages was also seen in the inter-lobar measures and the distribution features that described the extent of diffuse, intermediate and dense disease in the lung. Moderate to strong correlations were observed in most cases.

There were various limitations in the study. I did not have access to airway wall thickness data, which may be a significant feature in the prediction of exacerbation frequency from CT scans as seen by Han et al. [87]. Their work was performed on the COPDGene cohort. Inclusion of this feature in my analysis would have been important. In effect, they observed an interesting non-linear relationship between airway wall thickness, total emphysema and exacerbation frequency. Future work should include this feature to study how it varies with the presented global and regional measures of disease spread.

A further issue in the work presented relates to the definition of exacerbation susceptibility. Controversies remain over the exact definition of exacerbations [37]. Patients may only inform their healthcare practitioners only in the most severe cases [240]. As a result, it is possible that the exacerbation frequency data studied is not accurate. Moreover, the data

relates to past exacerbation episodes. It may be a theoretically flawed process to attempt to predict a clinical variable that defines past behaviour and not a current clinical state.

Lastly, there was possibly an interaction effect that was ignored in the experiments that would need to be included in future models (Section 6.3.3). Furthermore, the experiment design in Section 6.3.4 was also possibly flawed. The hypothesis was to test whether the addition of various new features aided in the prediction of exacerbation frequency. However, model performance was biased by the addition of many non-significant features and selection of the best model was not possible. A better strategy would be to employ Random Forests, which implicitly ranks features when building the decision trees. The most important features could then be used for further analysis.

6.5 Conclusion

In summary, I have presented a pipeline that facilitates the extraction of various CT-features that aim to capture global and regional properties of lung pathophysiology. I presented new features such as metrics that define diffuse and dense disease to those that capture interlobar variations of disease. I observed moderate to strong correlations between all features and FEV_1%predicted, suggesting that these should be employed in future studies when assessing links between CT, spirometry and other markers of COPD severity.

The presented pipeline was applied to the study of exacerbation susceptible COPD. This is the first study to my knowledge that has analysed global and regional CT imaging features in relation to exacerbation susceptible and non-susceptible COPD patients. This was also the first study that has investigated a potential link between various measures of lung deformation and the exacerbation susceptible phenotype. I did not find any significant relationships between the proposed features and exacerbation susceptibility. However, I observed that FEV_1%predicted is the strongest predictors of past exacerbations. Further work is necessary to evaluate whether the proposed measures are clinical useful in the analysis and prediction of COPD exacerbations.

Chapter 7

Event-based modelling and clustering of COPD

The work presented in this chapter is a joint project between F. Bragman and A. Young*.

∗ **Contributions**

I initiated the collaboration with Alexandra Young. We both designed the initial experiments, which involved selecting the features and choosing the number of clusters to model. Alexandra Young developed the SuStaIn model [249]. Her implementation was employed within this chapter and applied to the COPDGene data, which I supplied. I performed all the resulting analysis and drafting of this chapter.

Abstract

This thesis has been based on the development of tools to better study and diagnose Chronic Obstructive Pulmonary Disease (COPD) from Computed Tomography (CT). I have presented various techniques that aim to quantify and exploit the distribution and regional extent of CT features to study various clinical questions. In Chapter 3, I exploited the distribution of density-based and biomechanical features to classify exacerbation susceptible COPD. In Chapter 4, I presented a lobe segmentation algorithm that was applied to study regional disease extent in relation to COPD exacerbation susceptibility in Chapter 6. The issue with the analysis however is that it ignored potential patient-differences in the progression of COPD. The quantified CT features provide a snap-shot of the disease process and do not differentiate between different progression trajectories. The work I presented in 5 attempted to address this issue by building progression trajectories of emphysema and airway disease.

In this chapter, I present to my knowledge, the first application of disease progression modelling to the study of COPD. I used a model called SuStaIn (**S**ubtype and **S**tage **I**nference), that was developed by A. Young to study neurodegenerative diseases. I applied it to COPDGene, which is a large multi-centre study aimed at precisely phenotyping COPD using clinical, imaging and genetic data. The studied cohort was cross-sectional with subjects at various degrees of lung function impairment in addition to smoking controls. As a pilot study, I determined if clusters of patients with similar disease progression trajectories could be uncovered. I found two main data-driven subgroups of COPD with various associated clusters. Importantly, each cluster represented sets of patients with distinct disease progression patterns in contrast to traditional COPD clusters where patients have common clinical and CT measures. The disease progression models has also further strengthened the notion that early diagnosis of COPD is crucial to inhibit the acceleration of lung function impairment.

7.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is heterogeneous in the population with a large variability of potential clinical manifestations. The rate of FEV_1 decline amongst smokers is highly variable as seen in the ECLIPSE cohort over a 3-year longitudinal study [226]. Since alterations to lung tissue and the airway tree are the primary drivers of airflow limitation, this heterogeneity in lung function decline may be partly explained by varying contributions of emphysema and airway disease. It is also possible that FEV_1 may in fact be a poor clinical marker when compared to new multi-dimensional metrics such as the BODE (Body mass index, airway Obstruction, Dyspnoea and Exercise capacity) index [30].

There is a need for better clustering of COPD patients for precise, personalised medicine. Effective management of COPD requires clinicians to understand the natural progression of disease whilst appropriately managing therapeutic interventions with unique clinical phenotypes. Unsupervised machine learning algorithms techniques known as clustering can help with this task. Clustering of COPD has been performed by identifying emphysema subgroups [131, 21, 245] and through determining distinct groups of patients, with common clinical features and similar responses to therapy [133]. Applied to COPD, various subsets previously unknown have been identified [31, 27, 73]. However, a fundamental limitation of these techniques is that patients in identified clusters may potentially not have a similar temporal progression of disease. In fact, it is possible that patients within the same cluster are at different stages of distinct progression routes. The quantification of different progression routes in COPD may yield novel insights about specific phenotypes of COPD.

Data-driven disease progression modelling concerns a family of statistical models aimed at estimating the sequence of events (discrete) or the biomarker trajectory (continuous) that best characterise the progression of a disease from an early, asymptomatic stage to the most severe stage. Modelling patterns of disease progression may be critical for prognosis since it permits both an estimation of a patient's progress in the course of disease and a prediction of future progression based on a number of observed biomarker values. The development of disease progression modelling techniques have originated in the study of neurodegenerative diseases. A variety of fully data-driven techniques have been developed [250, 68, 55, 183, 102]. They facilitate the quantification of the full temporal resolution of disease and have helped provide novel insights into the neurodegenerative disease progres-

sion patterns, particularly in Alzheimer's Disease.

The key similarity, which makes applying these models to the study of COPD particularly attractive lies in the assumption that neurodegenerative diseases are not reversible. Akin to the progressive and irreversible nature of COPD, neurodegenerative diseases are chronic progressive disorders characterised by the loss of structure and function of neurons. Disease progression modelling of neurodegenerative diseases has shown that independent anatomical units of the brain are pathologically altered in a sequential fashion [68]. More recent developments are now able to cluster patients based on their sequence of events [250]. Despite the obvious links, application of these techniques to the study of COPD and other progressive respiratory diseases has been limited. Vogl et al. [229] developed a spatio-temporal disease pattern distribution model to monitor idiopathic pulmonary fibrosis (IPF) and classify patients into IPF with Usual Interstitial Pneumonia and IPF without Usual Interstitial Pneumonia. Applied to COPD, Boes et al. [25] derived a three-compartment model that simulates progression of tissue from normal, to functional small airways disease and emphysema. The ensuing set of differential equations governing the model can be used to estimate the progression of COPD.

The work presented in this thesis has focused on the development of methods to better quantify COPD phenotypes and severity from CT scans. However, as a multi-factorial disease, there are various potential mechanisms that act during the progression of disease. Despite two patients experiencing equal levels of severity for a specific biomarker e.g. emphysema extent in the lung, they may be at different stages in their own progression whilst being part of different phenotypes of disease. These two patients may in fact belong to distinct groups, or disease subgroups, with different progressions of disease caused by different underlying biological mechanisms. Application of the SuStaIn model [250] can tackle this problem by uncovering distinct subgroups of COPD with differing progression of disease.

In this chapter, I present the application a novel unsupervised machine learning technique developed by Young et al. [250] called SuStaIn (**S**ubtype and **S**tage **I**nference) to quantify the progression of COPD. It is based on the work of Fonteijn et al. [68] and allows for the simultaneous clustering of disease subgroups and characterisation of disease stage.

7.2 SuStaIn: Subtype and Stage Inference

7.2.1 Event-based model of disease progression

The SuStaIn model was built on the event-based disease progression model of Fonteijn et al. [68], which is described in this section. It is built on two main assumptions. The first is that the disease processes are irreversible and cannot revert back to initial states. This is mirrored in COPD, as it is clinically defined as a progressive, irreversible disease. The second assumption is that there is a common sequence of events within the studied cohort. This assumption was later removed by Young et al. [249] to allow for clustering of different event sequences.

The model consists of a series of events $E = \{E_i \,|\, i = 1, \cdots, N\}$ with an ordering $S = \big(s(1), \cdots, s(N)\big)$. The ordering S consists of a permutation of the numbers $i = 1, \cdots, N$, which determines the event ordering of E such that $E_{s(1)}, \cdots, E_{s(N)}$. The events are indexed by $i = 1, \cdots, N$ for all patients $j = 1, \cdots, J$. They are contained in the full data matrix $X \in \mathbb{R}^{J\,xN}$ where $X_j = \{x_{1j}, x_{2j}, \cdots, x_{Nj}\}$ corresponds to the measurements of patient j. Given a control population and a COPD population, both data matrices are defined as $X_{control}$ and X_{COPD}.

The events are specified before the modelling and correspond to specific biomarkers or variables that are hypothesised to be formative in the progression of disease. In the model, they are specified as either normal (healthy) or abnormal (pathological). In the context of COPD, these events could correspond to airway wall thickening, gas trapping, mucus hypersecretion and emphysema. The ordering S would therefore specify the sequence upon which these events occur in the progression of COPD.

In order to fit the disease progression model, evaluating the probability density function $P(S\,|\,X)$ is required. This measures the likelihood of a particular sequence S given the measurements of the cohort in X. The sequence S that maximises $P(S\,|\,X)$ is consequently the most likely ordering of events given the studied data.

To determine the most likely sequence, the likelihood that a particular measurement is still normal or has become abnormal must be quantified and is defined for both respectively as $P(x_{ij}\,|\,\neg E_i)$ and $P(x_{ij}\,|\,E_i)$. Both distributions are independent of j and can be quantified *a priori* with X_{COPD} and $X_{control}$. In the work of Fonteijn et al. [68], a mixture model with a Gaussian component ($P(x_{ij}\,|\,\neg E_i)$) and uniform distribution ($P(x_{ij}\,|\,E_i)$) is fit to X_{COPD} and $X_{control}$. Given a patient measurement x_{ij}, $P(x_{ij}\,|\,\neg E_i)$ and $P(x_{ij}\,|\,E_i)$ are easily quantified.

If a patient j is at position k in the sequence S, events $E_{s(1)},\cdots,E_{s(k)}$ have occurred whilst $E_{s(k+1)},\cdots,E_{s(N)}$ have not. The joint probability of observing the patient's data (X_j) given the sequence S whilst the patient is at event k can be formulated as

$$p(X_j\,|\,S,k)=\prod_{i=1}^{k}p(x_{ij}\,|\,E_{s(i)})\prod_{i=k+1}^{N}p(x_{ij}\,|\,\neg E_{s(i)}). \tag{7.1}$$

By assuming independence of the measurements across all patients and marginalising the hidden variable k, the total likelihood can be formulated as

$$p(X\,|\,S)=\prod_{j=1}^{J}\left[\sum_{k=0}^{N}p(k)\left(\prod_{i=1}^{k}p(x_{ij}\,|\,E_{s(i)})\prod_{i=k+1}^{N}p(x_{ij}\,|\,\neg E_{s(i)})\right)\right] \tag{7.2}$$

where $P(k)$ is a uniform prior such that $P(k)=1/N$. The probability density function $P(S\,|\,X)$ can be obtained by application of Bayes' theorem

$$P(S\,|\,X)=\frac{p(S)\,p(X\,|\,S)}{p(X)} \tag{7.3}$$

The above equation can be solved using a Markov Chain Monte Carlo algorithm (MCMC), which yields a set of T candidate orderings. The characteristic ordering $\bar{S}$, represents the most likely event sequence, which occurs most frequently in the set of candidates. This ordering thus represents the modelled biomarker trajectory of the studied cohort X_{COPD}.

7.2.2 The SuStaIn model

The model of Fonteijn et al. [68] assumes an instantaneous switch from normal to abnormal for a biomarker in the progression of disease. However, a biomarker is much more likely to follow a specific progression pattern as it evolves from normal to abnormal. This is the theory underlying the SuStaIn model [249]. In SuStaIn, a subtype is referred to as a group of subjects with a particular biomarker progression pattern. This biomarker progression is described as a series of events, where each event corresponds to a biomarker reaching a particular z-score compared to a control group. The process is modelled as a linear z-score event-based model, which represents the disease progression as the continuous, monotonic, linear accumulation of biomarkers from one z-score to another. The final progression model of a subgroup is defined as a mixture of linear z-score event-based models (Figure 7.1).

The second extension of SuStaIn with respect to the original event-based model is the

number of clusters. The SuStaIn model assumes a fixed number of subgroups $c = 1, \cdots, C$ in a studied cohort. This number can either be be pre-set or determined through cross-validation. The specific ordering S_c of each cluster and the position of the patient in S_c are quantified through SuStaIn.

The linear z-score model consists of a set of N z-score events $E_{(i,z)}$, which correspond to the linear increase of a biomarker $i = 1, \cdots, I$ to a z-score $z_{(i,r)} = z_{(i,1)}, \cdots, z_{(i,R_i)}$ (Figure 7.1). The trajectory of a specific biomarker i is modelled as a piecewise linear function $g(t)$ on an arbitrary time axis t. A z-score event $E_{(i,z)}$ occurs at each disease stage k starting at time $t_{start} = {}^{(k)}/_{(N+1)}$ and ending at $t_{end} = {}^{(k+1)}/_{(N+1)}$. The sequence of events S_c for a specific subtype is therefore modelled by the ordering of all z-score events $E_{(i,z)}$ for the biomarkers $i = 1, \cdots, I$.

The linear z-score model quantifies sequential events in the timeline of the disease progression and is not temporal. The time axis (t) of the progression is arbitrary. It initiates from $t = 0$ to $t = 1$ where $t = 1$ signifies conclusion of the disease progression. In Figure 7.1, a hypothetical model is illustrated. Whilst one can infer that biomarker 1 ($g_{i=1}(t)$) reaches z_2 before biomarker 2 ($g_{i=2}(t)$), it is not possible to extrapolate whether this occurs at a faster or slower temporal pace i.e. months versus years. However, I can differentiate biomarker trajectories that reach z-scores in succession as fast-acting versus those that occur more sporadically across t as long-acting.

Inclusion of the linear z-score model in the overall model likelihood (Equation 7.2) and removal of the normal and abnormal likelihoods leads to the following

$$p(X \mid S) = \prod_{j=1}^{J} \left[\sum_{k=0}^{N} \left(\int_{t=\frac{k}{N+1}}^{t=\frac{k+1}{N+1}} \left(P(t) \prod_{i=1}^{I} P(x_{ij} \mid t) \right) \partial t \right) \right] \tag{7.4}$$

where the integral term represents the joint probability of a particular linear z-score model given the biomarker measurements of patient j. The prior $P(t) = 1$ and is assumed to be uniform.

The overall model (M) likelihood for all clusters C is defined as

$$p(X \mid M) = \sum_{c=1}^{C} f_c \, P(X \mid S_c) \tag{7.5}$$

where f_c is the fraction of patients in cluster c and C is the number of subgroups.

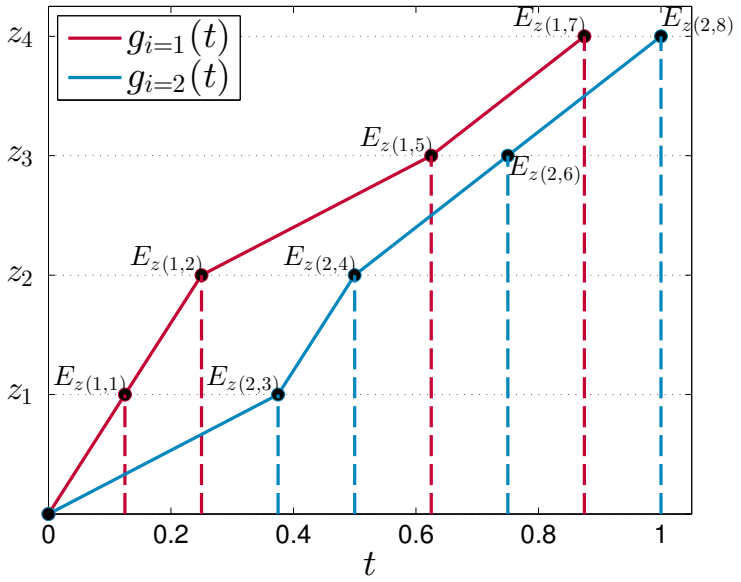


Figure 7.1: Linear z-score event-based model for a hypothetical model with $I = 2$ biomarkers and $N = 7$ events. The first event in this model is $E_{z(1,1)}$ and corresponds to biomarker $i = 1$ reaching a z-score equal to 1. The second event is $E_{z(1,2)}$, corresponding to $i = 1$ and $z = 2$. The second biomarker ($i = 2$) then becomes abnormal in the third event $E_{z(2,3)}$. The disease progression of this hypothetical model is thus defined by the series of events $E_{z(i,r)}$ govern by the biomarker progression $g_i(t)$.

A hierarchical method is employed to fit the SuStaIn model. Given the desired number of clusters (C_{max}), the clustering is performed sequentially from $C = 1$ to $C = C_{max}$. At each iteration, $C-1$ candidate sets of C clusters are generated by sequentially iterating through the $C-1$ clusters and finding optimal splits for each cluster. Each new candidate set is composed of the two new clusters (after optimal split) and the remaining $C-1$ clusters. These are used to initialise the model and generate the set of C clusters for each candidate set. Of the $C-1$ candidate sets, the model with highest likelihood is retained as the solution at iteration C. For example, at $C = 4$, there are 3 clusters (c_1, c_2, c_3) at the $C-1$ stage. Three candidate sets of clusters are produced by computing the optimal split of each c_1, c_2, c_3 e.g. candidate set 1 is equal to $\mathbb{C}_1 = \{c_{1a}, c_{1b}, c_2, c_3\}$, candidate set 2 is $\mathbb{C}_2 = \{c_1, c_{2a}, c_{2b}, c_3\}$ and candidate set 3 is $\mathbb{C}_3 = \{c_1, c_2, c_{3a}, c_{3b}\}$. SuStaIn is then run $C-1 = 3$ times using $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$ as initial conditions to the fitting. If $\mathbb{C}_1$ has the highest likelihood, the new set of clusters for $C = 4$ is $\mathbb{C}_1 = \{c_{1a} \rightarrow c_1, c_{1b} \rightarrow c_2, c_2 \rightarrow c_3, c_3 \rightarrow c_4\}$.

7.3 Experiments and results

I applied the SuStaIn model using data from the COPDGene study [167]. The aim was to perform unsupervised clustering using only routine measurements that can be acquired from Computed Tomography (CT) scans. Since each cluster represents a distinct group with a shared common biomarker trajectory, application of SuStaIn allowed for the first time the quantification of various progression patterns of COPD. Furthermore, the linear z-score model underlying the SuStaIn methodology facilitated the detection of subgroups with faster acting events versus longer-term periods of pathological damage in the timeline of the disease progression.

7.3.1 Data

7.3.1.1 COPDGene cohort

The COPDGene study [167] is one of the largest studies to date aiming to investigate the underlying genetic factors of COPD. To identify genetic risk factors and determine susceptibility to COPD and other related phenotypes, a total of 10,000 subjects were recruited. This included healthy smoking controls and patients across the spectrum of COPD severity (GOLD 1, 2, 3 and 4). There are two major components to the COPDGene study: 1) cross-sectional analysis and 2) long-term longitudinal follow-up of all patients. For the duration of the work presented in this thesis, only the cross-sectional component of the COPDGene study had been completed. No longitudinal follow-up data was available. The data supplied by COPDGene presented the unique opportunity to apply disease progression modelling and potentially uncover various subgroups from cross-sectional data that differ based on their disease progression. However, no validation of the progression models could be performed.

7.3.1.2 Patient selection

I downloaded the entire database of patient measurements from the COPDGene cohort [167]. This consisted of $10,092$ patients with 344 associated variables. There were $4,437$ COPD patients (GOLD ≥ 1) and $4,245$ smoking controls (GOLD$= 0$). The remainder ($1,347$) were unclassified such that they had a $^{\text{FEV}_1}/_{\text{FVC}} > 0.7$ but an FEV_1%predicted < 80 or they were non-smoking controls. From the $4,437$ patients with COPD and the $4,245$ controls, I only selected patients who had data available for all CT features of interest (Section 7.3.1.3). This led to a total number $J_{COPD} = 1,360$ and $J_{control} = 1,163$ used in the SuStaIn

model.

7.3.1.3 COPD features

A mixture of CT features from the COPDGene variable set were considered. This allowed me to account for different aspects of COPD pathology (Table 7.1). The first family of features (tissue) concerned those that can be detected by analysis of the voxel attenuation. The two most common densitometric features are emphysema and gas trapping. Emphysema is quantified by computing the percentage of voxels in the lung with an intensity below the threshold of -950HU on a scan at total lung capacity. A metric for gas trapping, which indirectly measures effects of small airways disease, is obtained by analysis of scans acquired at forced residual capacity. It is defined as the percentage of voxels with an attenuation below -856HU. The COPDGene database provided results for upper-lobe and lower-lobe emphysema. Given the plethora of work studying effects of apical and basal predominant emphysema, lobar quantities of emphysema were considered over whole lung values of emphysema.

The second family of features used was based on measurements of airway morphology. Airway abnormalities such as airway wall thickening and luminal narrowing are hallmarks of chronic bronchitis and bronchiolitis, which are both distinct phenotypes of COPD. Airway measurements from CT such as the Wall Area % (WA%) and the Pi10/15 square-root of wall area (SRWA) aim to capture these effects and were used in this study.

The WA% is the ratio between the wall area and total bronchial area. It measures the degree of airway thickening with respect to luminal area. The COPDGene database provided the WA% for segmental and sub-segmental airways. The segmental bronchi occur after division of the lobar bronchi and supply air to regions of the lung known as the bronchopulmonary segments. They are the fourth generation of the airway tree, averaging an airway diameter of 5mm to 8mm in a healthy lung [146]. The sub-segmental airways are in in the sixth and eighth-generation and have approximately a mean diameter of 3mm [146]. They lead to the respiratory bronchioles where gas exchange is initiated.

To measure pathology associated to the lower parts of the respiratory tree, the Pi10 and Pi15 square-root of wall area (SRWA) measures have emerged [148]. Measurements are taken in the large and intermediate-sized airways, which can be accurately assessed using CT in order to predict the wall thickness of the smaller airways. The SRWA is measured in the upper-zones of the airway tree. By measuring the perimeter of these airways (Pi), the

typical SRWA of airways with a 10mm (3.18mm diameter) and 15mm (4.78mm diameter) lumen perimeter can be predicted by a linear regression. It has been seen that dimensions of the large and intermediate-sized airways reflect the dimensions of smaller airways [148].

7.3.1.4 Quantitative CT analysis

The data used in the study (Table 7.1) was processed by COPDGene and downloaded from the database. The analysed CT-features were obtained by analysis of paired breath-hold CT scans by COPDGene collaborators using the VIDA Pulmonary Workstation [1].

Airway dimensions for segmental and subsegmental airways were collected for six bronchial paths: 1) right upper lobe apical bronchus (RB1), 2) right middle lobe posterior basal bronchus (RB4), 3) right lower lobe posterior basal bronchus (RB10), 4) left upper lobe apicoposterior bronchus (LB1), 5) superior lingular bronchus (LB4) and 6) left lower lobe posterior basal bronchus (LB10). These were chosen based on the consensus of COPDGene investigators [166]. Airway indices were measured from the centreline of the airways to the airway edge in each slice of the image volume. Measures for wall area percentage (WA%) were quantified for each branch. For each patient, measurements are averaged for the segmental and subsegmental data. Moreover, measures for Pi10 SRWA and Pi15 SRWA were determined through the measurements of all these branches.

Table 7.1: Variables used to model COPD progression with the SuStaIn model. SRWA - square root of wall area, WA - wall area, A_o - total area and A_i - luminal area.

Type	Feature	Quantification
Tissue	Upper-lobe emphysema	$\% < -950$HU in the upper lobes at inspiration
	Lower-lobe emphysema	$\% < -950$HU in the lower lobes at inspiration
	Whole lung gas trapping	$\% < -856$HU in the expiration scan
Airway	Pi10 SRWA	Prediction of $\sqrt{(}$WA$)$
	Pi15 SRWA	
	Wall Area % segmental	$100 \times (A_o - A_i)/(A_o)$
	Wall Area % sub-segmental	

7.3.2 COPD progression models

The SuStaIn model was applied by seeking out two (Model 1) and five clusters (Model 2) of COPD. The optimal number of clusters given the studied cohort X_{COPD} should be determined by cross-validation prior to analysis. However, the aim of this experiment was to firstly, dichotomise the COPDGene cohort into its two main disease progression patterns

[1] vidadiagnostics.com

(Model 1) then further analyse what clusters could be quantified with SuStaIn by increasing the number of clusters (Model 2). The number of clusters for Model 2 was then increased to five to allow SuStaIn to uncover more progression trajectories.

7.3.2.1 Model 1: two-cluster model

The biomarker trajectories for both clusters can be visualised in Figure 7.3. The proportion of GOLD classified patients in both clusters is shown in Figure 7.2. A χ^2 test of independence was performed by comparing the frequency of GOLD stages between both clusters. I found significant differences between clusters ($\chi^2(DF = 3) = 7.82, p < 10^{-5}$). Cluster 1 had significantly more GOLD 1 and 4 patients whilst GOLD 2 and 3 were more expressed in cluster 2.

There are important differences between both clusters, which highlight two potential main phenotypes of COPD (Figure 7.3). These can be categorised by the sequence of events and the linear accumulation of the z-scores. Cluster 1 can be classified as a Tissue-Airway progression pattern since features related to the voxel intensity become abnormal first. Conversely, cluster 2 is an Airway-Tissue progression pattern since pathological variations in the airway-related metrics are the first to become apparent.

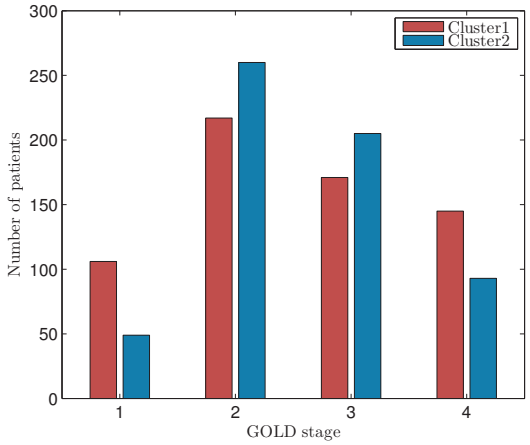
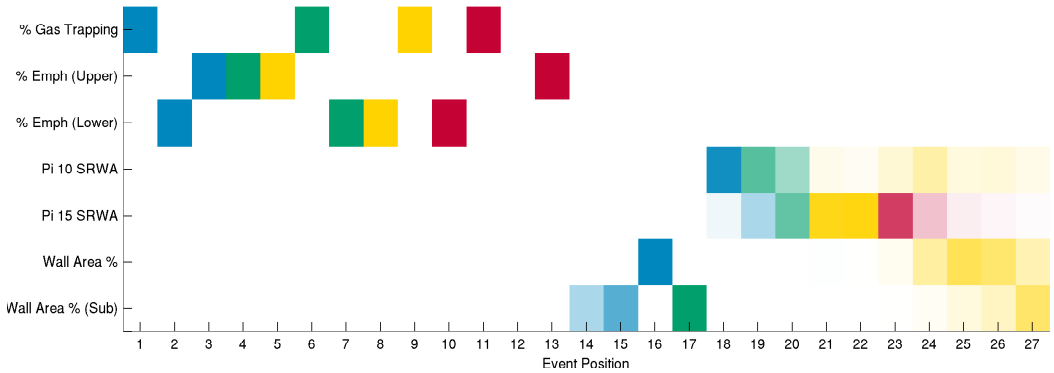
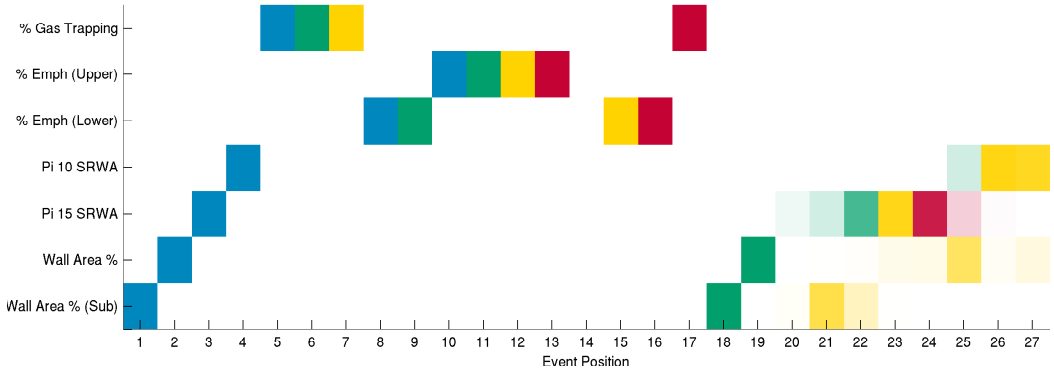


Figure 7.2: GOLD stage proportion in Model 1

In cluster 1 (Figure 7.3a), there are early signs of gas trapping then lower-lobe emphysema. This is followed by a rapid progression of upper-lobe emphysema, which is illustrated by the progression of z-score events. Gas trapping and lower-lobe emphysema then progress with lower-lobe emphysema approaching maximum severity first. Importantly, gas trapping and measures of emphysema progress to their most severe stage before initiation of the airway-related pathology. These are manifested first by airway wall remodelling in



(a) Cluster 1 - 51% of COPDGene cohort



(b) Cluster 2 - 49% of COPDGene cohort

Figure 7.3: Model 1 - two cluster disease progression model. Opacity of the z-score events quantifies uncertainty in the event ordering.

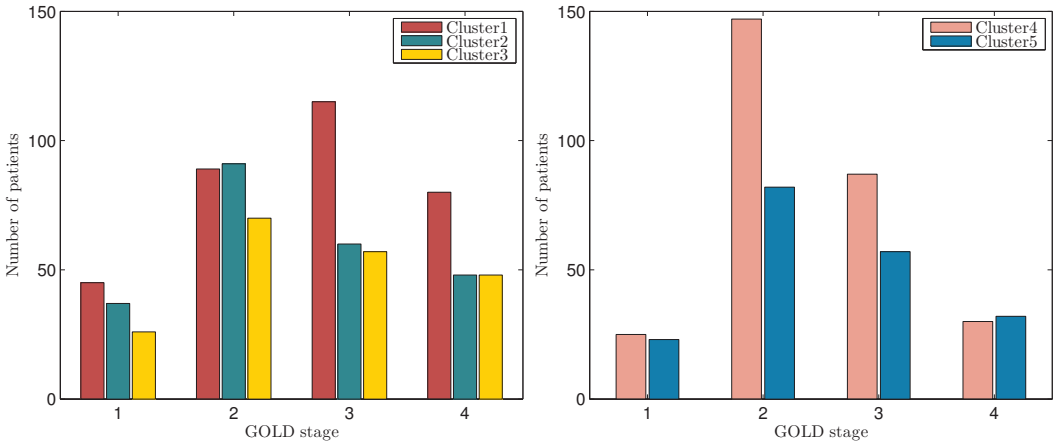
the subsegmental then segmental airways.

The progression of cluster 2 (Figure 7.3b) is initiated by a cascade of events in the airways. Abnormalities are first present in airway walls of subsegmental then segmental airways. This is closely followed by changes in the Pi15 SRWA and Pi10SRWA features. However, the progression of these features appears to be long-lasting over the spectrum of events. Pathological alterations in the airways directly lead to gas trapping, which is fast-acting and aggressive within this cluster. This is then followed by a slow progression of emphysema in the lower lobes and fast acting emphysema in the upper lobes.

7.3.2.2 Model 2: five-cluster model

An increase in the amount of clusters yielded a new set of biomarker trajectories with distinct progression patterns (Figure 7.5 and 7.6). I observed a dichotomy in the population based on the sequences of events. There were three Tissue-Airway clusters (Figure 7.5)

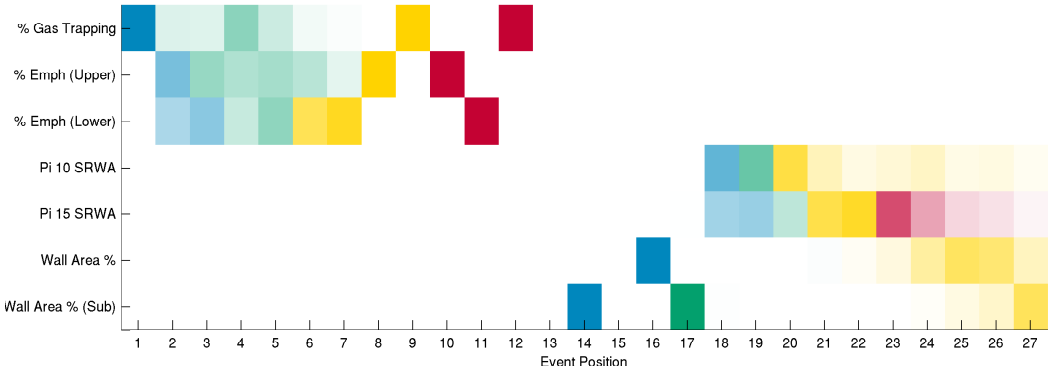
and two Airway-Tissue clusters (Figure 7.6). A χ^2 test of independence within the Tissue-Airway and the Airway-Tissue subgroups was performed respectively to test for differences in the GOLD proportion of subjects (Figure 7.4). There was no significant difference ($\chi^2(DF = 6) = 12.12, p = 0.06$) between the clusters within the Tissue-Airway subgroup (Figure 7.4a). There was also no significant difference ($\chi^2(DF = 3) = 6.41, p = 0.09$) within the Airway-Tissue subgroup (Figure 7.4b).



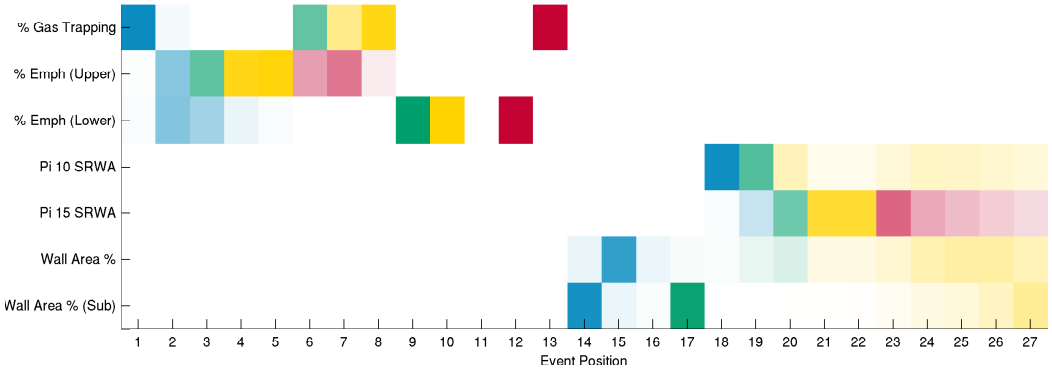
(a) Tissue-Airway subgroups **(b)** Airway-Tissue subgroups

Figure 7.4: GOLD stage proportion in Model 2

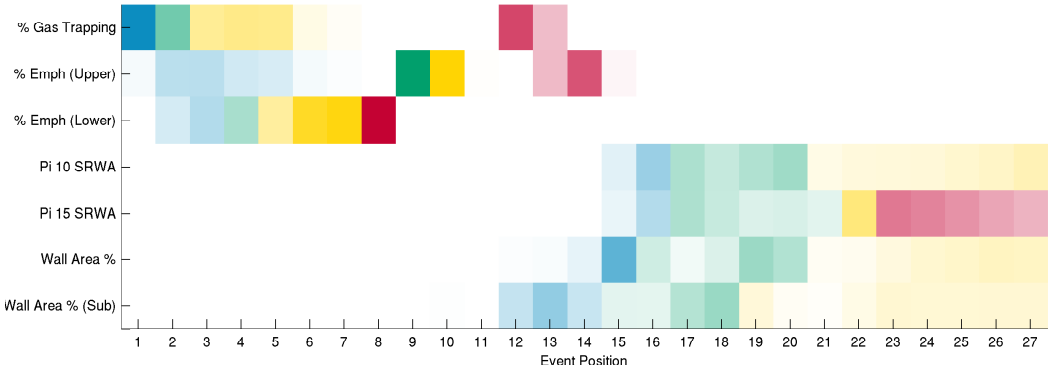
The biomarker trajectories of the Tissue-Airway clusters (Table 7.2, Figure 7.5a, 7.5b and 7.5c) can be characterised by the z-score accumulation and whether events occur synchronously. In the first cluster (Figure 7.5a), progression of gas trapping in addition to emphysema in the upper and lower lobes occurs synchronously. Once these biomarkers reach maximum severity, airway-related pathologies are initiated. Airway wall thickening of the sub-segmental airways first occurs, which is followed by thickening of the segmental region, and followed by a synchronous progression of all biomarkers until conclusion of the disease progression. The progression of airway-related pathology was similar in clusters 2 (Figure 7.5b) and 3 (Figure 7.5c). The most distinct differences were in the progression of upper and lower lobe emphysema. Cluster 2 can be defined as a rapid upper-lobe emphysema sequence (Figure 7.5b). It reaches the highest severity of upper-lobe emphysema at event 7 in contrast to event 10 and 14 in clusters 1 and 3 respectively. The opposite was observed in cluster 3 (Figure 7.5c), which can be described as an aggressive lower-lobe emphysema subgroup. The presence of gas trapping is first apparent then followed by a rapid increase in lower lobe emphysema.



(a) Cluster 1 - 33% of COPDGene cohort



(b) Cluster 2 - 19% of COPDGene cohort



(c) Cluster 3 - 13% of COPDGene cohort

Figure 7.5: Model 2 - tissue-airways clusters. Opacity of the z-score events quantifies uncertainty in the event ordering.

Table 7.2: Characteristics of the Tissue-Airway clusters

Model	Biomarker	Property
Cluster 1 - Figure 7.5a	Gas trapping Upper-lobe emphysema Lower-lobe emphysema	Synchronous progression
Cluster 2 - Figure 7.5b	Gas trapping Upper-lobe emphysema Lower-lobe emphysema	Fast-acting upper-lobe emphysema
Cluster 3 - Figure 7.5c	Gas trapping Upper-lobe emphysema Lower-lobe emphysema	Fast-acting lower-lobe emphysema

The SuStaIn model quantified two clusters that were determined to be part of the Airway-Tissue family (Table 7.3, Figure 7.6a and 7.6b). In the first cluster (Figure 7.6a), the progression of COPD is initiated by abnormalities in wall area percentage in the segmental and sub-segmental airways (WA%). This is then followed by a rapid progression of the Pi10 and Pi15 SRWA metrics, which suggested that remodelling of the airways occurs at an aggressive pace in this COPD subgroup. A sequence of rapid events in the tissue related features occurs after the initial progression of airway-related metrics. Gas trapping abnormalities reach a severe z-score rapidly followed by a synchronous worsening of emphysema in the upper and lower lobes.

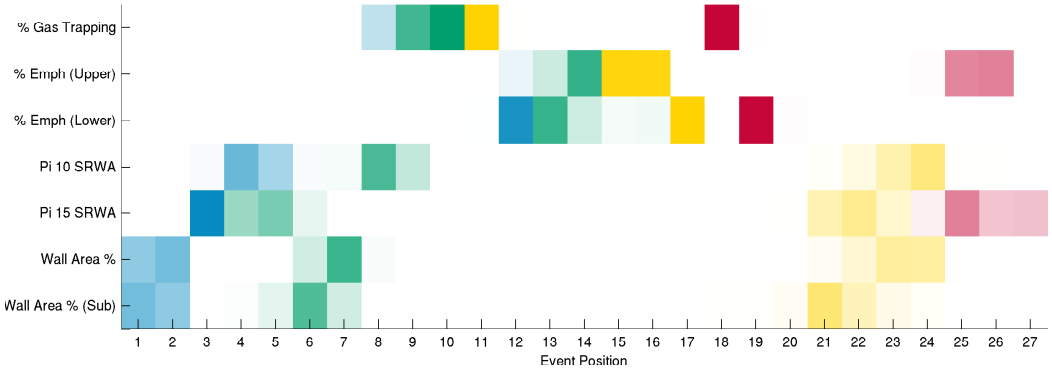
A more uncertain initiation in the progression of the second Airway-Tissue cluster was observed (Figure 7.6b). Gas trapping and abnormalities in wall area percentage initiates the progression pattern. A domino effect in the sequence of tissue-related events is then seen. Gas trapping first progresses to a moderate severity stage, followed by the rapid progression of upper-lobe emphysema and longer-acting lower-lobe emphysema.

7.3.3 COPD disease severity and disease progression

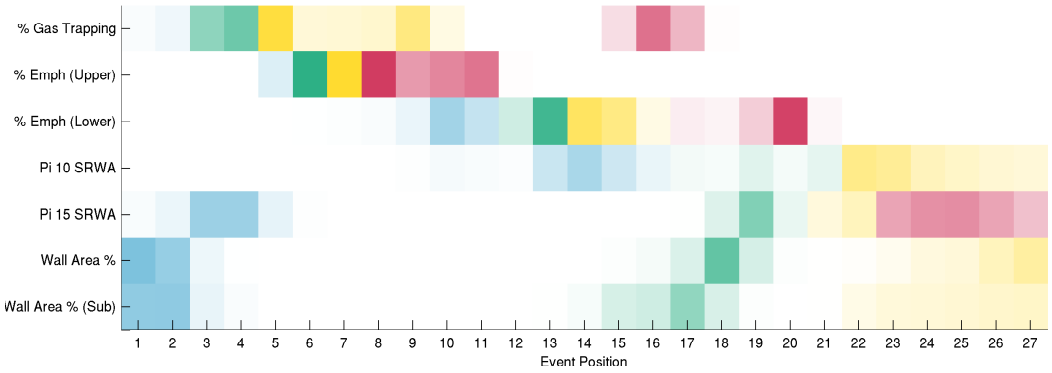
Application of the SuStaIn model to a large population of COPD patients allowed me to divide the COPDGene cohort into two major subgroups of disease; Tissue-Airway and Airway-Tissue models of COPD progression. Both these models contained distinct clusters, which are characterised by the event sequences and the aggressiveness of specific biomarker progressions (Table 7.2 and 7.3). An important property of SuStaIn is that each patient was indexed within the clusters by their respective progression stage E. This allowed me to test

Table 7.3: Characteristics of the Airway-Tissue clusters

Model	Biomarker	Property
Cluster 4 Figure 7.6a	Pi10 SRWA Pi15 SRWA Segmental WA% Sub-segmental Segmental WA%	Fast-acting airway and tissue pathology
Cluster 5 Figure 7.6b	Pi10 SRWA Pi15 SRWA Segmental WA% Sub-segmental Segmental WA%	Slow-acting airway pathology



(a) Cluster 4 - 24% of COPDGene cohort



(b) Cluster 5 - 12% of COPDGene cohort

Figure 7.6: Model 2 - airway-tissue clusters. Opacity of the z-score events quantifies uncertainty in the event ordering.

whether COPD severity, as defined by the disease progression event E was consistent with the staging of severity through spirometry (Section 7.3.3.1) but also exacerbation frequency (Section 7.3.3.3). This analysis was performed on model 2 (5 clusters).

7.3.3.1 Disease progression event stage and GOLD classification

Associations were first investigated between the progression event stage and GOLD classification (Table 7.4) in the 5-cluster model (Section 7.3.2.2). Significant differences were found across the Tissue-Airway progression clusters for each GOLD stage yet none in the Airway-Tissue subgroup (Table 7.4).

A post hoc Tukey-Kramer test was performed for all three clusters of the Tissue-Airway subgroup at each GOLD stage. A consistent significant increase was found in the mean event of patients in cluster 1 with respect to cluster 3 except at GOLD 4 (GOLD 1: $p = 0.03$, GOLD 2: $p < 10^{-3}$, GOLD 3: $p < 10^{-5}$, GOLD 4: $p = 0.21$). There was only a significant increase between clusters 2 and 3 at GOLD 3 ($p < 10^{-}5$). There were no other significant differences across clusters for the event stage. It can be seen that clusters 1 and 2 of the Tissue-Airway subgroup have on average, consistently higher mean event stages per GOLD classification. This suggests a more rapid deterioration of the lung anatomy before lung function becomes abnormal in clusters 1 and 2.

Table 7.4: Mean (St.D.) of progression stages for of all patients per GOLD stage stratified by cluster

Subgroup	Cluster	GOLD 1	GOLD 2	GOLD 3	GOLD 4
Tissue-Airway	1	5.91 (3.41)	7.88 (3.45)	11.54 (3.03)	14.62 (2.99)
	2	4.35 (2.58)	6.84 (3.50)	12.45 (3.56)	14.69 (3.50)
	3	3.96 (3.13)	5.73 (3.81)	8.95 (4.43)	13.46 (5.12)
One-way ANOVA		$p = 0.01$	$p < 10^{-3}$	$p < 10^{-5}$	$p = 0.19$
Airway-Tissue	4	2.88 (1.67)	5.82 (4.07)	10.30 (5.19)	16.27 (4.56)
	5	3.74 (2.14)	5.61 (3.78)	11.16 (4.71)	14.22 (4.78)
One-way ANOVA		$p = 0.13$	$p = 0.70$	$p = 0.32$	$p = 0.09$

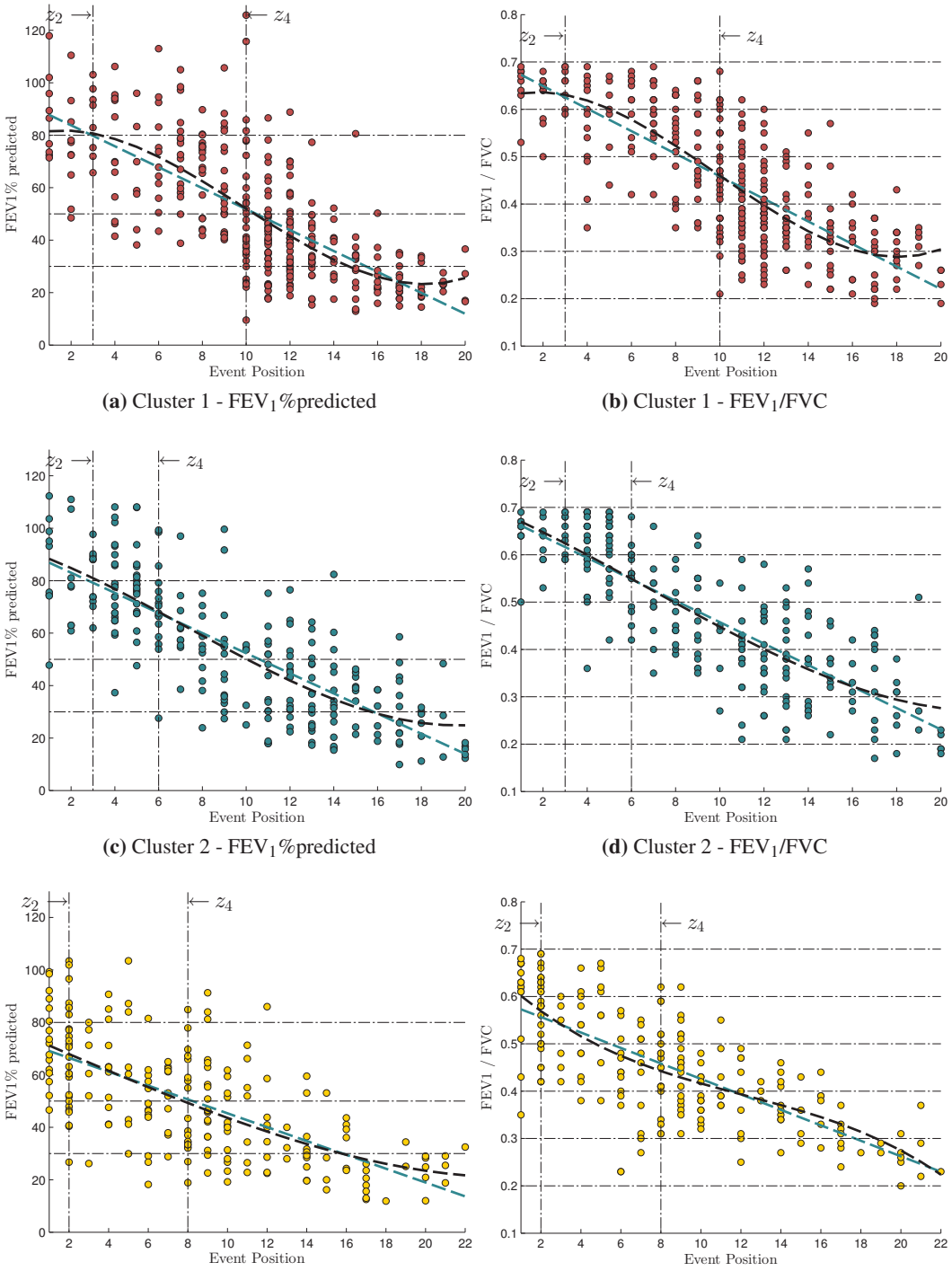
7.3.3.2 Disease progression event stage and spirometry

Relationships between event stage and spirometric indices (FEV_1%predicted and FEV_1/FVC) were examined in each cluster. For each cluster, the event stage of the patient was plotted versus their respective spirometric values (Figure 7.7 and 7.8). Pearson's Linear Correlation was calculated between the event stage and spirometry. Strong significant correlations were observed across clusters (Table 7.5).

I also determined the event stage upon which biomarkers first became moderately abnormal (z_2) and very severe (z_4) (Table 7.5). For instance, within cluster 1 of the Tissue-Airway subgroup (Figure 7.5a), lower-lobe emphysema first reached z_2 at event 3 and upper-lobe emphysema was the first biomarker to reach z_4 at event 10. These were overlaid in Figures 7.7 and 7.8 as vertical lines. This illustrated cases within each cluster where patients had advanced severity as defined by the event stage yet mild to moderate COPD as seen by spirometry. Various sets of patients in each subgroup were observed with discordant relationships between their disease progression stage and their respective lung function score. For example, cluster 2 (Figure 7.5b) and cluster 5 (Figure 7.6b) had various patients with at least one biomarker at the most severe (z_4) yet with mild lung function impairment.

Table 7.5: Event stage for feature abnormality and Pearson's Linear Correlation (ρ) of event stage with spirometry. [* : $p < 10^{-5}$]

Subgroup	Cluster	z-score	Feature	$\rho(\text{FEV}_1\%\text{p})$	$\rho\left(\text{FEV}_1/\text{FVC}\right)$
Tissue-Airway	1	$z_2 = 3$ $z_4 = 10$	Lower-lobe emphysema Upper-lobe emphysema	-0.72^*	-0.76^*
	2	$z_2 = 3$ $z_4 = 6$	Upper-lobe emphysema Upper-lobe emphysema	-0.77^*	-0.81^*
	3	$z_2 = 2$ $z_4 = 8$	Gas trapping Lower-lobe emphysema	-0.65^*	-0.73^*
Airway-Tissue	4	$z_2 = 4$ $z_4 = 18$	Pi15 SRWA Gas trapping	-0.66^*	-0.72^*
	5	$z_2 = 3$ $z_4 = 8$	Gas trapping Upper-lobe emphysema	-0.70^*	-0.73^*



(a) Cluster 1 - FEV_1%predicted

(b) Cluster 1 - FEV_1/FVC

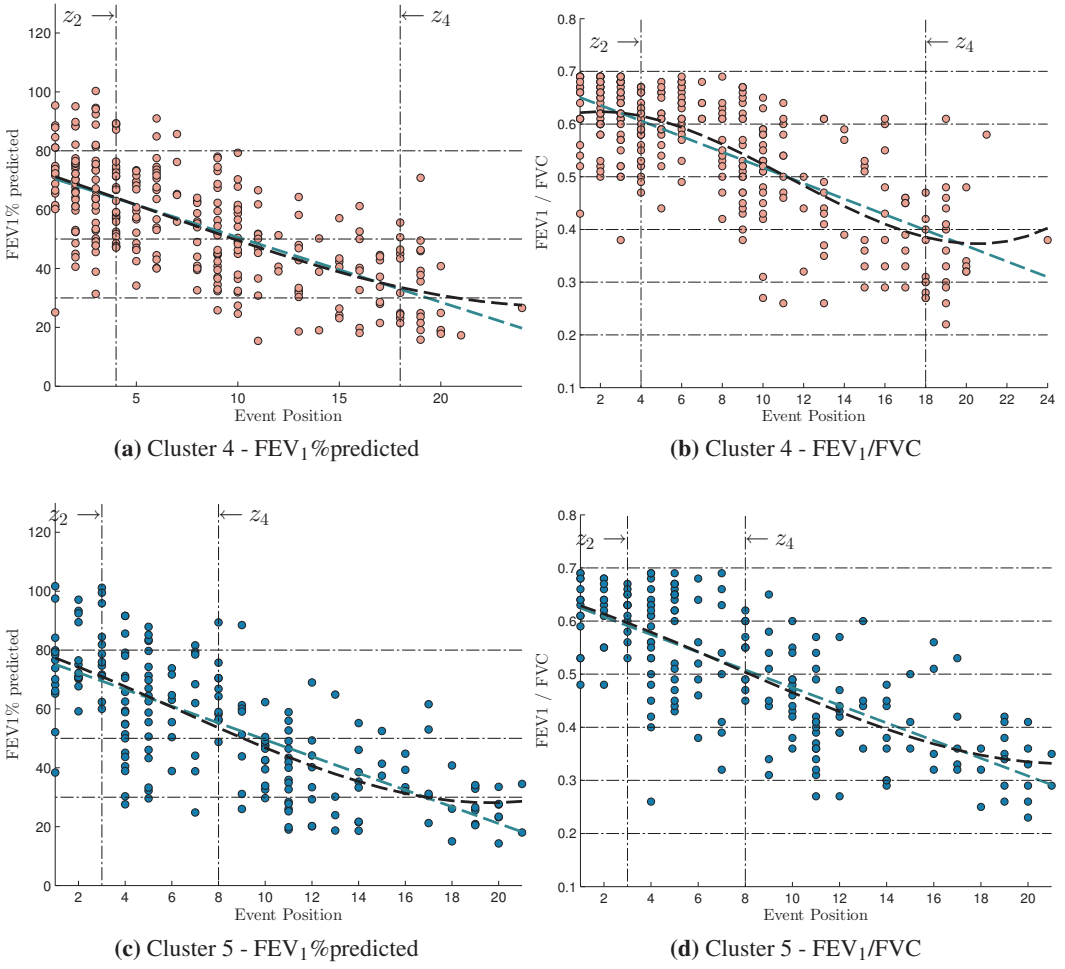
(c) Cluster 2 - FEV_1%predicted

(d) Cluster 2 - FEV_1/FVC

(e) Cluster 3 - FEV_1%predicted

(f) Cluster 3 - FEV_1/FVC

Figure 7.7: Relationship between disease progression and spirometry in the Tissue-Airway subgroup. The horizontal lines in the FEV_1% predicted plot correspond to the thresholds used in GOLD classification for GOLD 1, 2, 3 and 4. Those used in the FEV_1/FVC plots were arbitrarily chosen.



(a) Cluster 4 - FEV_1%predicted

(b) Cluster 4 - FEV_1/FVC

(c) Cluster 5 - FEV_1%predicted

(d) Cluster 5 - FEV_1/FVC

Figure 7.8: Relationship between disease progression and spirometry in the Airway-Tissue subgroup. The horizontal lines in the FEV_1% predicted plot correspond to the thresholds used in GOLD classification for GOLD 1, 2, 3 and 4. Those used in the FEV_1/FVC plots were arbitrarily chosen.

7.3.3.3 Exacerbation susceptibility and COPD disease progression

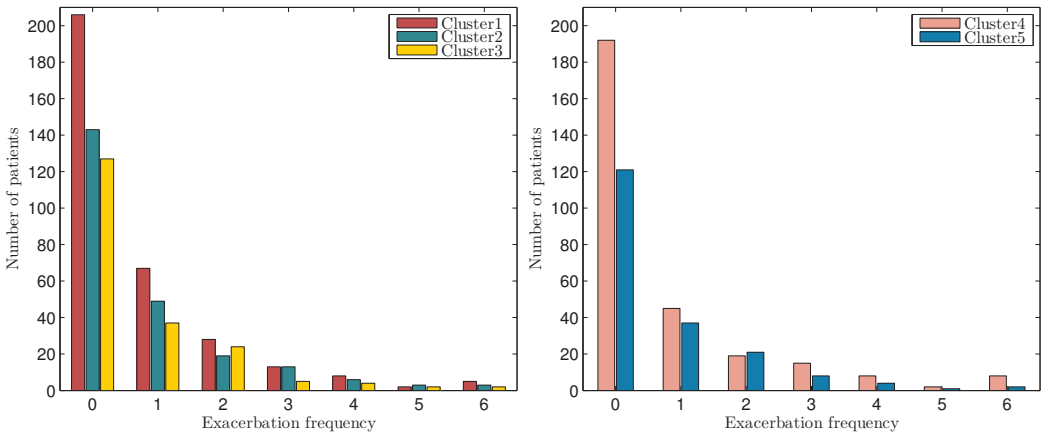
Various studies have focused on testing associations between CT features and exacerbation frequency [44, 87, 203]. However, none of these studies have been able to test effectively whether there is a relationship between disease progression of COPD and susceptibility to exacerbations of COPD. Therefore, I investigated whether there were any relationships between the clusters and exacerbation susceptibility. I also analysed correlations between disease event stage within the clusters and exacerbation frequency.

The frequency of susceptible ($f \geq 2$) and non-susceptible ($f < 2$) COPD patients in each subgroup and cluster (Figure 7.9 and Table 7.6) was first analysed. I first tested whether the subgroups had different distributions of susceptible and non-susceptible patients. They

both had the same frequency of patients (Table 7.6) and this was corroborated in a χ^2 test of independence ($\chi^2(DF = 1) = 0.02, p = 0.88$). There was no significant difference in the ratio of susceptible to non-susceptible patients between both subgroups; suggesting that there is no leading progression of disease for the susceptible subgroup of COPD.

I also tested the hypothesis that there were differences between the distribution of exacerbation frequencies between subgroup clusters (Figure 7.9a and 7.9b). I found no significant difference between the three clusters in the tissue-airway subgroup ($\chi^2(DF = 12) = 6.17, p = 0.91$) and also within the airway-tissue subgroup ($\chi^2(DF = 6) = 5.93, p = 0.43$). In line with our results assessing differences between the tissue-airway and airway-tissue subgroups, these results confirmed no specific within subgroup differences in the disease progression of susceptible and non-susceptible patients.

Furthermore, I calculated Spearman's rank order correlation between exacerbation frequency and disease progression event stage. I observed weak relationships in all tissue-airway clusters and also within cluster 5 of the airway-tissue subgroup (Table 7.7).



(a) Tissue-Airway subgroup **(b)** Airway-Tissue subgroup

Figure 7.9: Frequency of exacerbations for each COPD progression cluster

7.4 Discussion

A novel unsupervised machine learning algorithm has been applied for the first time to COPD imaging data. The SuStaIn model has identified two dominant subgroups of COPD; the Tissue-Airway and Airway-Tissue subgroups. Within the Tissue-Airway subgroup, the presence of gas trapping and emphysema first became apparent, which was followed by airway-related abnormalities. In the Airway-Tissue subgroup, disease progression was ini-

Table 7.6: Frequency of exacerbation susceptible and non-susceptible patients per COPD clusters

Subgroup	Cluster	Susceptible	Non-Susceptible	χ^2-within	χ^2-between
Tissue-Airway	1	0.17	0.83	$p = 0.91$	
	2	0.18	0.82		
	3	0.19	0.81		
	All	0.18	0.82		
					$p = 0.88$
Airway-Tissue	4	0.18	0.82	$p = 0.43$	
	5	0.19	0.81		
	All	0.18	0.82		

Table 7.7: Spearman's rank-order correlation between exacerbation frequency and disease progression event stage per COPD cluster

Subgroup	Cluster	ρ	p-value
Tissue-Airway	1	0.21	$< 10^{-}5$
	2	0.27	$< 10^{-}5$
	3	0.22	0.002
Airway-Tissue	4	0.09	0.14
	5	0.31	$< 10^{-5}$

tiated by airway-related pathologies.

The clustering and staging of SuStaIn has facilitated the quantification of distinct clusters within the Tissue-Airway and Airway-Tissue COPD subgroups (Figure 7.5 and 7.6, Table 7.2 and 7.3). By analysing the event stage of patients and their respective spirometric readings, I have been able to show that significant structural alterations in lung tissue can occur prior to these changes becoming apparent through spirometry (Figure 7.7 and 7.8). The speed of progression of the biomarker trajectories also differed across clusters and subgroups, which may suggest different underlying mechanisms responsible for the disease process. This discordance between disease event position and clinical measures of lung function are also indicative of the need to detect early the progression of COPD before the bulk effect of pathology initiates a rapid progression in lung function decline.

I also investigated potential relationships between exacerbation susceptibility and the quantified clusters. Patients that are susceptible to COPD exacerbations are an important high-risk group. It is not known if the disease progression of susceptible patients differs from the non-susceptible group. I have observed that there is no correlation between disease progression and susceptibility to exacerbations. This may suggest that the mechanism

responsible for the progression of COPD is distinct from the underlying cause of exacerbation susceptibility. Whilst I have found no significant differences in the susceptibility to exacerbations between clusters, isolating COPD patients from large cohorts for focused within-cluster analysis may yield key differences in contrast to a cohort-wide study. This could be applied to analysis performed in Chapter 6. Moreover, future work should aim to include exacerbation frequency in the SuStaIn modelling to determine whether there are distinct clusters where exacerbations occur prior to CT structural alterations and vice-versa.

The results confirm in part findings from previous unsupervised clustering efforts [32] on COPDGene. Castaldi et al. [32] identified four separate COPD clusters by using CT features and spirometric measures such as functional residual capacity (FRC) % predicted, FEV_1% predicted and bronchodilator responsiveness as a percentage of FEV_1. The identified clusters that are similar to those observed in this chapter are the: 1) mild-upper zone predominant emphysema with airflow obstruction cluster and 2) airway predominant disease cluster. A key difference is that the identified clusters in my work represent patients with similar progression routes of disease and patients can be at various different stages of their progression. In contrast, clustered patients in the work of Castaldi et al. [32] may possibly be following distinct progression routes yet may have similar features. Future work, aimed at bridging the clustering of Castaldi et al. [32] with the SuStaIn subgroups will be able to effectively determine this.

There is a strong case for the early detection of COPD. The work of Fletcher and Peto [67] initially outlined the rate of lung function decline in COPD; suggesting slow decline at onset followed by a more rapid phase at advanced stages of disease. Recent studies have suggested that faster progression of lung functional impairment occurs earlier [207] and particularly at mild-to-moderate stages of COPD [207, 58]. These results are mirrored in studies showing that smokers may develop emphysema, measured in CT before exhibiting abnormal lung function [181, 195]. These undetected structural alterations may be critical in the early, accelerated decline of lung function and the remaining course of COPD progression. The results from Section 7.3.3.2, which showed relationships between disease progression event stage and spirometry provide evidence of the need to early detect COPD progression. Patients may already be at stages of advanced disease yet at mild-to-moderate stages of COPD. This is illustrated by the thresholds z_2 and z_4 in Figures 7.7 and 7.8 and summarised in Table 7.5.

Despite the novel findings and potential exciting future avenues of research stemming from this chapter, there is a crucial need to validate the disease progression patterns. I found strong correlations between event stage and spirometry (Section 7.3.3.2), which signifies that the progression trajectories model the continuum of lung function impairment from moderate to very severe. Morever, the two uncovered subgroups (Tissue-Airway and Airway-Tissue) reflect the current knowledge in COPD phenotypes (emphysema versus airway predominant) [200], which weakly validates the results. However, longitudinal data is necessary to perform robust validation of the results. The data supplied and made available from COPDGene at the time of this project was cross-sectional. This was not sufficient to validate the modelled longitudinal trajectories. Existing frameworks that exist for simulating cross-sectional or longitudinal biomarker datasets from disease cohorts [251] could be used to evaluate the performance of the modelling presented in this chapter. However, the follow-up longitudinal data that will be released as part of the COPDGene study will help reflect the accuracy of the predictions made in the modelled subgroups. Whilst the timescale of COPD progression is over decades, it is expected that employing the follow-up data and validating on separate data sources such as the ECLIPSE study [225] will help elucidate whether the uncovered clusters are true subtypes of COPD and help determine the clinical utility of this modelling.

7.5 Conclusion

In conclusion, disease progression modelling of COPD has allowed me to quantify two main subgroups of COPD; Tissue-Airway and Airway-Tissue. It also allowed me to uncover clusters within each subgroup, which differed by the aggressiveness of the biomarker changes and the sequence upon which they became abnormal. These results may have great clinical repercussions as they suggest that patients can be stratified by their most likely progression of disease. Whilst the analysis must be repeated by determining the optimal number of clusters and the clusters must be validated using independent and longitudinal data, they may be important starting-points for assessing potential genetic associations and differences in therapeutic response. This has the potential to revolutionise the management of COPD and improve future treatment of patients.

Chapter 8

Summary and discussion

This aim of this thesis has been to develop automated tools to analyse Computed Tomography (CT) scans to improve the analysis and diagnosis of Chronic Obstructive Pulmonary Disease (COPD). There are various components of COPD that cause pathological alterations in a number of pulmonary structures, which in combination lead to airflow limitation and other comorbidities. The method upon which COPD affects the lung may also vary widely across patients such as in upper-lobe and lower-lobe predominant disease. There are consequently various subtypes of COPD with potentially important clinical differences. This necessitates precise phenotypic characterisation of COPD subjects from CT.

In this chapter, I summarise the work presented within this thesis. I present an overview of the main original contributions of each chapter. I then suggest the future directions of this project and initial results when present.

8.1 Summary

8.1.1 Classification of exacerbation susceptible COPD

Chapter 3 presented a framework for classifying exacerbation susceptible and non-susceptible patients although it could be applied to staging COPD severity. This could be performed in a multi-class setting by predicting patient GOLD stage. The spread of disease and its effect throughout the lung can vary from homogeneity to heterogeneity. Such complex changes cannot be accurately represented by classification tools that exploit average measures [143, 23]. Moreover, patients can be susceptible or non-susceptible to exacerbations at equal levels of functional impairment with similar levels of emphysema. This inhibits the application of algorithms that classify patients based on local probabilities of COPD [194, 193] or on dissimilarity measures of lung texture [192].

These limitations were addressed by attempting to quantify a signature unique to the patient that is based on the distribution of density-based and biomechanical features across the lung. This was performed by analysing features locally and then modelling their global distribution. This distribution could be exploited after dimensionality reduction using Principal Component Analysis as features in a classifier. I demonstrated the performance of the algorithm by classifying exacerbation susceptible and non-susceptible patients at equal levels of COPD severity (GOLD 3). This was to my knowledge the first time this had been performed. Despite the complexity of the problem, I showed that it is possible to classify exacerbation susceptible patients at equal levels of severity although the performance of the algorithm was variable when tested on various hold-out sets in a repeated k-fold cross-validation. Recent work however has attempted to classify exacerbation frequency using convolutional neural networks with promising results [248].

The future application of the algorithm lies in the analysis of the distributions with respect to COPD severity. Before attempting challenging problems such as the classification of susceptible patients, it is necessary to gain a better understanding of how the distributions vary in a population of control and COPD patients.

A potentially important aspect of the framework is sampling the local distributions and their ensuing global distribution. This may implicitly capture information about emphysema subtypes. Previous work has classified emphysema subtypes based on local ROI intensity distributions [131] whilst fibrotic disease patterns can be disentangled from emphysema using a PCA on integral geometry features [39]. More recent work has exploited texton-based, difference of Gaussian (DOG) and local binary patterns (LBP) features for unsupervised labelling of ROIs [245]. Parameterisation of local intensity distributions with the addition of textural features and their ensuing distribution across the lung may model the spread of varying emphysema subtypes.

8.1.2 Segmentation of the pulmonary lobes

In Chapter 4, I presented a lobe segmentation algorithm. The regional quantification of disease may be critical in monitoring disease progression but also in patient stratification for surgery. It has been seen that a more homogeneous presentation of emphysema accelerates decline of FEV_1 [206]. The presence of apical or basal emphysema may have discordant effects on lung physiology [81, 75]. Identifying patients that have upper-lobe predominant emphysema is also important for staging lung volume reduction surgery [36].

The lobe segmentation algorithm combined the segmentation of the fissures with a groupwise fissure prior to find segmentation labels for the lobes. The main aim of the algorithm was for it to be unsupervised at all stages of the pipeline. The state of the art in fissure segmentation is based on supervised classification [223], hard-coded empirical parameters [109, 238], shape models [174] and line-enhancing filters [242]. I presented a method to segment the fissures by assuming a generative model between a multi-scale fissure enhancement filter and the underlying segmentation of the lung. It can be argued that the method is not fully unsupervised since it requires setting parameters for the filter. I showed that the method was generally insensitive to the parameter choice. The method could be made fully unsupervised by merging the segmentation of fissures obtained from a range of filter parameters through majority voting. I also demonstrated the utility of constructing a groupwise fissure prior from a set of complete and incomplete fissure segmentations without prior need for manual segmentations. Constructing a deformable model of the fissures without manually labelled fissures is difficult. However, I showed that its application had the ability to guide the segmentation in regions of incomplete fissures. The algorithm achieved state of the art performance when tested on multiple datasets. However, there were still outliers in the performance of the algorithm notably due to errors in the airway seed labelling. In the future, the dependence on the airway tree for seed labelling should be reduced.

8.1.3 Manifold learning of COPD

I presented a new method to quantify the spread of emphysema and functional small airways disease (fSAD) in Chapter 5. Current methods for quantifying disease extent in the lung rely on global averages and lobar averages. They may not be able to capture local aspects of lung disease and quantify their spread. This makes it difficult to accurately quantify inter-patient differences. Sørensen et al. [192] developed a new distance metric that relies on bipartite-graph matching of local ROIs. Harmouche et al. [88] presented a new severity measure based on the extent of various emphysema subtypes within the lung. A fractal exponent can also also be derived, which represents the size distribution of emphysematous clusters [134]. Despite these advances, techniques that can accurately quantify inter-patient differences in disease spread within pulmonary image analysis are limited and sparse.

I introduced the concept of local disease and deformation distributions as methods to better phenotype COPD from CT. This extended the framework introduced in Chapter 3. By sampling tissue classification results locally throughout the lung, a disease distributions

can be built, which captures the local properties of disease spread and how it varies globally across the lung. I also built a deformation distribution, which quantified local volume change across the lung. Inter-patient differences could be quantified by considering histogram distances. I applied a framework previously employed to study neonatal brain development [4] to investigate the progression of the proposed distributions. Manifold learning enabled me to capture variations in the distributions in a large cohort of patients. Fusion of the embeddings facilitated the construction of a single embedding, which encapsulated various aspects of COPD. Analysis of the embedding coordinates showed that the presented distributions associated more strongly with common markers of COPD severity than conventional metrics. The position of a patient in the manifold space of COPD may be an instrumental factor in the severity of disease. Importantly, kernel regression in the manifold space allowed me to build potential trajectories of disease progression. These however need to be validated. Longitudinal data that will be released as part of the COPDGene study [167] may be instrumental in this validation. This data was not available at the duration of this project. However, a more interesting future application lies in merging the results with the disease progression modelling in Chapter 7. This will demonstrate if the distance travelled along the modelled trajectories represents a viable metric of disease progression.

8.1.4 From global to regional analysis of lung CT: application to exacerbation susceptible COPD

In Chapter 6, I exploited the lobe segmentation algorithm (Chapter 4) and the disease distributions (Chapter 5) to develop a pipeline that analyses lung CT from a global to a regional level. The aim was to extract features that measure different aspects of lung disease, which can be used to investigate various hypotheses related to COPD. The pipeline was applied to the study of exacerbation susceptible COPD to investigate whether regional information associates with exacerbation frequency. Studies have currently shown that airway wall thickness associates with exacerbation frequency yet the relationship with emphysema is highly non-linear [87] with contrasting findings in the literature [203, 44, 69]. It is not known if fSAD associates with exacerbation frequency, whether regional differences differentiate susceptible and non-susceptible patients and whether lung deformation varies with exacerbation frequency.

I showed with the pipeline that many new features associate with FEV_1%predicted. Inter-lobar measures of emphysema, lung deformation in addition to measures of diffuse,

intermediate and dense emphysema and fSAD were all strongly correlated with spirometry. Novel CT-metrics such as Jacobian heterogeneity are being developed to investigate new links between parenchymal destruction and lung function impairment [18]. Furthermore, they may improve the prediction of respiratory morbidity and mortality [22]. Further studies are needed to assess whether the features employed in the analysis significantly improve the link between CT and COPD severity.

The analysis pipeline was able to find group-differences between the exacerbation susceptible and non-susceptible groups. Regional analysis was able to disentangle potential differences where global measures could not such as global fSAD, lower-lobe fSAD and inter-lobar fSAD. This suggested regional analysis of lung disease may be important for future studies. Univariate tests demonstrated that many of the features do associate with exacerbation frequency yet do not when FEV_1%predicted was included in the regression models. However, when accounting for all features in the most complex model, it was seen that volume change homogeneity and inter-lobar fSAD may be predictive of exacerbation frequency. These features may be worthy of future studies to evaluate their potential link with exacerbations.

8.1.5 Disease progression modelling of COPD

Chapter 7 presented the application of disease progression modelling to evaluate COPD disease progression. I applied SuStaIn (**S**ubtype and **S**tage **I**nference) developed by Young et al. [249], which is a novel unsupervised algorithm that bridges clustering with event-based modelling. Event-based modelling concerns a set of statistical models that aim to estimate the sequence of events that best characterise the progression of disease from an early, asymptomatic stage to the most severe stage. The SuStaIn model is able to cluster patients based on the trajectory of disease progression whilst also finding their respective stage in their potential course of disease.

I applied SuStaIn to 1360 COPD patients from COPDGene [167]. I was able to uncover two potentially important subtypes of COPD; Tissue-Airway and Airway-Tissue subtypes. Within the Tissue-Airway subtype, disease progression was initiated by various abnormalities in lung tissue then followed by alterations to the airway tree. Airway abnormalities first became apparent in the Airway-Tissue subtype, which led to the progression of tissue related pathology. There were various clusters within each subtype with differing progressions of disease. By analysing disease stage with respective spirometric readings,

I was able to reinforce the notion that early detection in COPD is critical as significant pathology may be present before spirometry becomes abnormal. This clustering approach when applied to COPD is truly novel since previous clustering efforts have grouped patients based on cross-sectional data [232, 33, 27, 73]. Clustering COPD patients based on likely disease progression has the potential for better patient stratification for focused analysis in large clinical studies such as COPDGene [167] but importantly, to help better understand our understand of the general history of COPD.

8.2 Future directions

The work presented within this thesis has focused on the development of quantitative tools for the study and diagnosis of COPD. The underlying motivation has been to develop a suite of techniques, which facilitates the precise phenotypic characterisation of COPD from CT. There are two main future directions arising from my work, which should be pursued. The first is centered on extending the work of Chapter 3, 5 and 4 towards a statistical framework to create models of health and disease. The second relies on extending Chapter 7 by including exacerbation susceptibility and frequency in the mathematical framework of SuStaIn whilst including the features presented in Chapter 6.

8.2.1 Statistical models of lung structure and deformation

The lung anatomy is incredibly complex and varies widely across the population. Coupled with the discussed complexities of COPD, it is obvious that there is a great need to define a normative range of pulmonary variation to clarify subject-invariant structure-function relationships and better identify deviations from health to disease. Statistical modelling in the context of pulmonary image analysis is however not a new concept. Computational models of respiratory motion have been studied extensively in the field of radiotherapy [128]. Statistical models of lung motion have been built to predict lung tumour motion [60, 116]. Groupwise analysis based on the construction of an atlas has been employed for voxel-wise regression analysis to evaluate response to treatment in Hyperpolarised-MRI [213]. These techniques can be exploited to better study COPD.

There are two main challenges that need to be addressed to create statistical models of lung structure and function. Accurate inter-patient groupwise registration is necessary and methods for analysing textural and deformation-based information in the groupwise space need to be developed.

Inter-patient lung registration was performed in Chapter 4 to create the groupwise fissure prior and represents a good starting point for future work. In order to improve the accuracy of the fit, anatomical landmarks can be employed. This was previously performed by Li et al. [112, 113] who developed a hybrid landmark and intensity-driven registration algorithm and manually identified corresponding points on the airway tree. Identifying corresponding landmarks is laborious and smarter techniques for introducing landmarks are possible. The fissure segmentation presented in Chapter 4 could be used to help drive the registration. This was performed by Schmidt-Richberg et al. [182]. Other options include applying the Frangi vesselness filter [172] or the vessel density [54] simultaneously with the original intensity image in a multi-channel registration.

There are many interesting applications that can be pursued if I consider that an accurate groupwise space has been achieved. A Principal Component Analysis (PCA) could first be applied on the resampled intensity images. The principal components may correlate with various clinical markers of lung function impairment whilst the principal modes of variations could be used to visualise regions of the lung that correlate most strongly by extending the analysis to voxel-based morphometry [9]. This sees a direct application in studies that are attempting to ascertain how regional variations in emphysema affect lung function. More advanced dimensionality reduction can also be used such as manifold learning, which was applied in Chapter 5. This would negate the need for constructing local disease distributions for inter-patient quantification.

The framework developed by Ehrhardt et al. [60] can also be applied to analyse lung deformation in the groupwise space. The deformation field obtained from the paired breath-hold registration of each patient can be mapped into the common groupwise space. This may help determine normative values of lung deformation that are normalised by the patient anatomy in a control population to understand the deformation of lung that has been affected by COPD.

Statistical modelling of lung structure and function may be instrumental in phenotyping COPD from CT. A major question within this thesis has been centered on exacerbation susceptible COPD; does an exacerbation susceptible lung differ significantly from a non-susceptible lung at equal levels of severity? By constructing separate statistical models of exacerbation susceptible and non-susceptible lungs, our understanding of the potential differences may be improved, which may significantly impact future COPD treatment.

8.2.2 Disease progression modelling and susceptibility to exacerbations

The work presented in Chapter 7 applied event-based modelling to uncover various distinct progression trajectories of COPD. As a post-hoc analysis, I investigated potential relationships between the progression trajectories and susceptibility to exacerbations. I observed no specific cluster with significantly more susceptible patients. However, it is important to include exacerbation frequency as a biomarker in the modelling. A fundamental question in COPD regards the origin of exacerbations and progression of COPD. Is the presence of exacerbations a function of the progression of COPD and its inherent severity or do exacerbations act as a precursor to the pathological changes observed in the lung? Mechanisms by which exacerbations contribute to the decline of health status in COPD also remains largely unknown [170]. The preliminary analysis performed in Chapter 7 could not determine this. Inclusion of exacerbation variables as part of the modelling may potentially elucidate the relationship between exacerbation and COPD progression.

SuStaIn is based on the linear z-score accumulation of various biomarkers. It models the trajectory of each biomarker as it progresses from a normal to severe state through a z-score based on the deviation from the control population. Several changes need to be made to SuStaIn to allow it to effectively model continuous data with count data (exacerbation frequency, n/yr) and binary data (exacerbation susceptibility and exacerbation severity).

Lastly, the features developed in Chapter 6 provided regional information relating to emphysema, fSAD, lung deformation in addition to metrics defining diffuse and dense disease. These features can all be included in the modelling to produce a more fine-grained picture of the disease process. This inclusion however is not trivial. A hypothesis in the literature is that fSAD precedes emphysema in the progression of disease [71, 25, 98]. A fundamental question in the modelling must be answered. A strong assumption in the event-based model is that there is a monotonic increase in the biomarker measurements. If the hypothesis that fSAD precedes emphysema is correct, then it is expected that the extent of fSAD will decrease with disease progression as emphysema dominates. This violates the assumption of the model. Correct modelling of fSAD and emphysema progression in addition to the inclusion of exacerbations in the SuStaIn modelling are likely to have a significant impact in improving our understand of COPD.

8.3 Conclusion

This thesis has presented new automated tools to analyse Computed Tomography (CT) scans to improve the analysis and diagnosis of Chronic Obstructive Pulmonary Disease (COPD). The work presented has been motivated by the complexities of COPD and the need to better phenotype COPD from CT. My work has led towards a new method that classifies CT scans based on the distribution of local features. I have presented a new method to quantify disease extent in the lung and have shown a potential way to identify trajectories of disease progression from CT analysis. I have developed a novel lobe segmentation algorithm and have applied it to study whether regional analysis of lung disease associates with exacerbation susceptibility. Lastly, I demonstrated the possibility of identifying novel subtypes of COPD using event-based modelling.

List of publications

Papers in international journals

- **F. Bragman**, J. McClelland, J. Jacob, J. Hurst and D. Hawkes (2017). "Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior". In: *IEEE Transactions on Medical Imaging*, vol. 36 (8), pp 1650-1663.

Papers in conference proceedings

- **F. Bragman**, J. McClelland, M. Modat, S. Ourselin, J. Hurst and D. Hawkes (2014). "Multi-scale Analysis of Imaging Features and its Use in the Study of COPD Exacerbation Susceptible Phenotype". In: *Medical Image Computing and Computer Assisted Interventions - MICCAI 2014.* Edited by P. Golland, N. Hata, C. Barrillot, J. Hornegger and R. Howe. Lecture Notes in Computer Science, vol 8675, pp. 417-424.

- **F. Bragman**, J. McClelland, J. Jacob, J. Hurst and D. Hawkes (2017). "Manifold Learning of COPD". In: *Medical Image Computing and Computer Assisted Interventions - MICCAI 207.* Edited by M. Descoteaux, L. Maier-Hein, A. Franz, P. Collins and S. Duchesne. Lecture Notes in Computer Science, vol 10435, pp. 586-593.

Published abstracts

- **F. Bragman**, J. McClelland, D. Hawkes and J. Hurst (2014). "Multi-scale analysis of imaging features to study exacerbation susceptible COPD". In: *European Respiratory Society, Annual Congress, Munich.*

Bibliography

[1] R. Achanta, A. Shaji, and K. Smith. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2281, 2012.

[2] A. Agustí, P. Sobradillo, and B. Celli. Addressing the complexity of chronic obstructive pulmonary disease: From phenotypes and biomarkers to scale-free networks, systems biology, and P4 medicine. *American Journal of Respiratory and Critical Care Medicine*, 183(9):1129–1137, 2011.

[3] D. F. Alamidi, A. R. Morgan, P. L. Hubbard Cristinacce, L. H. Nordenmark, P. D. Hockings, K. M. Lagerstrand, S. S. Young, J. H. Naish, J. C. Waterton, N. C. Maguire, L. E. Olsson, and G. J. M. Parker. COPD Patients Have Short Lung Magnetic Resonance T1 Relaxation Time. *COPD*, 13(2):153–159, 2016.

[4] P. Aljabar, R. Wolz, L. Srinivasan, S. J. Counsell, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert. A combined manifold learning analysis of shape and appearance to characterize neonatal brain development. *IEEE transactions on medical imaging*, 30(12):2072–86, 2011.

[5] R. Amelon, K. Cao, K. Ding, G. E. Christensen, J. Reinhardt, and M. L. Raghavan. Three-dimensional characterization of regional lung deformation. *Journal of biomechanical engineering*, 44(13):2489–2495, 2011.

[6] American Thoracic Society. The definition of emphysema: report of a National Heart, Lung, and Blood Institute, Division of Lung Diseases Workshop. *American review of respiratory diseases*, 132:182–185, 1985.

[7] N. R. Anthonisen. Effects of smoking intervention and the use of an inhaled anti-

cholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA: The Journal of the American Medical Association*, 272(19):1497–1505, 1994.

[8] J. Ashburner. *Computational neuroanatomy*. PhD thesis, 2000.

[9] J. Ashburner and K. J. Friston. Voxel-Based MorphometryThe Methods. *NeuroImage*, 11(6):805–821, 2000.

[10] A. Aziz, K. Ashizawa, K. Nagaoki, and K. Hayashi. High resolution CT anatomy of the pulmonary fissures. *Journal of thoracic imaging*, 19(3):186–191, 2004.

[11] A. a. Bankier, C. R. O'Donnell, V. M. Mai, P. Storey, V. De Maertelaer, R. R. Edelman, and Q. Chen. Impact of lung volume on MR signal intensity changes of the lung parenchyma. *Journal of magnetic resonance imaging*, 20(6):961–6, 2004.

[12] P. J. Barnes. Chronic Obstructive Pulmonary Disease. *New England journal of medicine*, 343:269–280, 2000.

[13] M. M. Barreto, P. P. Rafful, R. S. Rodrigues, G. Zanetti, B. Hochhegger, A. S. Souza, M. D. Guimarães, and E. Marchiori. Correlation between computed tomographic and magnetic resonance imaging findings of parenchymal lung diseases. *European Journal of Radiology*, 82(9):492–501, 2013.

[14] N. K. Batmanghelich, A. Saeedi, M. Cho, R. S. J. Estepar, and P. Golland. Generative method to discover genetically driven image biomarkers. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9123, pages 30–42. Springer, Cham, 2015.

[15] G. Bauman, M. Puderbach, M. Deimling, V. Jellus, C. Chefd'hotel, J. Dinkel, C. Hintze, H. U. Kauczor, and L. R. Schad. Non-contrast-enhanced perfusion and ventilation assessment of the human lung by means of Fourier decomposition in proton MRI. *Magnetic Resonance in Medicine*, 62(3):656–664, 2009.

[16] A. Berrington de González. Projected Cancer Risks From Computed Tomographic Scans Performed in the United States in 2007. *Archives of Internal Medicine*, 169(22):2071, 2009.

[17] S. Bhatt, X. Soler, X. Wang, S. Murray, A. R. Anzueto, T. H. Beaty, A. M. Boriek, R. Casaburi, G. J. Criner, A. A. Diaz, M. T. Dransfield, D. Curran-Everett, C. Galban,

E. A. Hoffman, J. C. Hogg, E. A. Kazerooni, V. Kim, G. L. Kinney, A. Lagstein, D. A. Lynch, B. J. Make, F. J. Martinez, J. W. Ramsdell, R. Reddy, B. Ross, H. Rossiter, R. M. Steiner, M. Strand, E. J. R. Van Beek, E. Wan, G. R. Washko, J. M. Wells, C. Wendt, R. A. Wise, E. K. Silverman, J. D. Crapo, R. P. Bowler, and M. K. Han. Association between functional small airways disease and FEV 1 decline in COPD. *American journal of respiratory and critical care medicine*, 1164(e-press):201511–2219O, 2016.

[18] S. P. Bhatt, S. Bodduluri, J. D. Newell, E. A. Hoffman, J. C. Sieren, M. K. Han, M. T. Dransfield, and J. M. Reinhardt. CT-derived Biomechanical Metrics Improve Agreement Between Spirometry and Emphysema. *Academic Radiology*, (3), 2016.

[19] S. P. Bhatt, X. Soler, X. Wang, S. Murray, A. R. Anzueto, T. H. Beaty, A. M. Boriek, R. Casaburi, G. J. Criner, A. A. Diaz, M. T. Dransfield, D. Curran-Everett, C. Galban, E. A. Hoffman, J. C. Hogg, E. A. Kazerooni, V. Kim, G. L. Kinney, A. Lagstein, D. A. Lynch, B. J. Make, F. J. Martinez, J. W. Ramsdell, R. Reddy, B. Ross, H. Rossiter, R. M. Steiner, M. Strand, E. J. R. Van Beek, E. Wan, G. R. Washko, J. M. Wells, C. Wendt, R. A. Wise, E. K. Silverman, J. D. Crapo, R. P. Bowler, and M. K. Han. Association between functional small airways disease and FEV 1 decline in COPD. *American journal of respiratory and critical care medicine*, 1164(e-press):201511–2219O, 2016.

[20] J. Biederer, S. Mirsadraee, M. Beer, F. Molinari, C. Hintze, G. Bauman, M. Both, E. J. R. van Beek, J. Wild, and M. Puderbach. MRI of the lung (3/3)-current applications and future perspectives. *Insights into Imaging*, 3(4):373–386, 2012.

[21] P. Binder, N. K. Batmanghelich, R. S. J. Estepar, and P. Golland. Unsupervised discovery of emphysema subtypes in a large clinical cohort. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10019, pages 180–187. Springer, Cham, 2016.

[22] S. Bodduluri, S. P. Bhatt, E. A. Hoffman, J. D. Newell, C. H. Martinez, M. T. Dransfield, M. K. Han, J. M. Reinhardt, and COPDGene Investigators. Biomechanical CT metrics are associated with patient outcomes in COPD. *Thorax*, 72:409–414, 2017.

[23] S. Bodduluri, J. D. Newell, E. A. Hoffman, and J. M. Reinhardt. Registration-based lung mechanical analysis of chronic obstructive pulmonary disease (COPD) using a supervised machine learning framework. *Academic radiology*, 20(5):527–36, 2013.

[24] J. L. Boes, M. Bule, B. A. Hoff, R. Chamberlain, D. A. Lynch, J. Stojanovska, F. J. Martinez, M. K. Han, E. A. Kazerooni, B. D. Ross, and C. J. Galbán. The Impact of Sources of Variability on Parametric Response Mapping of Lung CT Scans. *Tomography*, 1(1):69–77, 2015.

[25] J. L. Boes, B. a. Hoff, M. Bule, T. D. Johnson, A. Rehemtulla, R. Chamberlain, E. a. Hoffman, E. a. Kazerooni, F. J. Martinez, M. K. Han, B. D. Ross, and C. J. Galbán. Parametric Response Mapping Monitors Temporal Changes on Lung CT Scans in the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS). *Academic Radiology*, 22(2):186–194, 2015.

[26] British Thoracic Society. The burden of lung disease: A statistics report from the British Thoracic Society. Technical report, 2006.

[27] P.-R. Burgel, J.-L. Paillasseur, D. Caillaud, I. Tillie-Leblond, P. Chanez, R. Escamilla, I. Court-Fortune, T. Perez, P. Carré, and N. Roche. Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *The European respiratory journal*, 36(3):531–9, 2010.

[28] M. J. Cardoso, M. J. Clarkson, G. R. Ridgway, M. Modat, N. C. Fox, and S. Ourselin. LoAd: A locally adaptive cortical segmentation algorithm. *NeuroImage*, 56(3):1386–1397, 2011.

[29] R. Casaburi, D. A. Mahler, P. W. Jones, A. Wanner, G. San Pedro, R. L. ZuWallack, S. S. Menjoge, C. W. Serby, and J. Witek. A long-term evaluation of once-daily inhaled tiotropium in chronic obstructive pulmonary disease. *European Respiratory Journal*, 19(2):217–224, 2002.

[30] C. Casanova, A. Aguirre-Jaime, J. P. De Torres, V. Pinto-Plata, R. Baz, J. M. Marin, M. Divo, E. Cordoba, S. Basaldua, C. Cote, and B. R. Celli. Longitudinal assessment in COPD patients: Multidimensional variability and outcomes. *European Respiratory Journal*, 43(3):745–753, 2014.

[31] P. J. Castaldi. Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax*, pages 1–8, 2014.

[32] P. J. Castaldi, J. Dy, J. Ross, Y. Chang, G. R. Washko, D. Curran-Everett, A. Williams, D. A. Lynch, B. J. Make, J. D. Crapo, R. P. Bowler, E. A. Regan, J. E. Hokanson, G. L. Kinney, M. K. Han, X. Soler, J. W. Ramsdell, R. G. Barr, M. Foreman, E. van Beek, R. Casaburi, G. J. Criner, S. M. Lutz, S. I. Rennard, S. Santorico, F. C. Sciurba, D. L. DeMeo, C. P. Hersh, E. K. Silverman, and M. H. Cho. Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax*, 69(5):415–22, 2014.

[33] P. J. Castaldi, J. Dy, J. Ross, Y. Chang, G. R. Washko, D. Curran-Everett, A. Williams, D. A. Lynch, B. J. Make, J. D. Crapo, R. P. Bowler, E. A. Regan, J. E. Hokanson, G. L. Kinney, M. K. Han, X. Soler, J. W. Ramsdell, R. G. Barr, M. Foreman, E. van Beek, R. Casaburi, G. J. Criner, S. M. Lutz, S. I. Rennard, S. Santorico, F. C. Sciurba, D. L. Demeo, C. P. Hersh, E. K. Silverman, and M. H. Cho. Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax*, 69(5):415–422, 2014.

[34] P. J. Castaldi, R. San José Estépar, C. S. Mendoza, C. P. Hersh, N. Laird, J. D. Crapo, D. a. Lynch, E. K. Silverman, and G. R. Washko. Distinct quantitative computed tomography emphysema patterns are associated with physiology and function in smokers. *American journal of respiratory and critical care medicine*, 188(9):1083–1090, 2013.

[35] R. Castillo, E. Castillo, D. Fuentes, M. Ahmad, A. M. Wood, M. S. Ludwig, and T. Guerrero. A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. *Physics in medicine and biology*, 58(9):2861–77, 2013.

[36] K. Cederlund, U. Tylén, L. Jorfeldt, and P. Aspelin. Classification of emphysema in candidates for lung volume reduction surgery: A new objective and surgically oriented model for describing CT severity and heterogeneity. *Chest*, 122(2):590–596, 2002.

[37] B. R. Celli and P. J. Barnes. Exacerbations of chronic obstructive pulmonary disease. *European Respiratory Journal*, 29(6):1224–1238, 2007.

[38] J. P. Charbonnier, E. M. Rikxoort, A. A. Setio, C. M. Schaefer-Prokop, B. van Ginneken, and F. Ciompi. Improving airway segmentation in computed tomography using leak detection with convolutional networks. *Medical Image Analysis*, 36:52–60, 2017.

[39] M. Charemza and A. Bhalerao. Integral geometry descriptors for characterizing emphysema and lung fibrosis in HRCT images. In *First International Workshop on Pulmonary Image Analysis*, New York, 2008.

[40] X. J. Chen, L. W. Hedlund, H. E. Möller, M. S. Chawla, R. R. Maronpot, and G. A. Johnson. Detection of emphysema in rat lungs by using magnetic resonance measurements of 3He diffusion. *Proceedings of the National Academy of Sciences of the United States of America*, 97(21):11478–81, 2000.

[41] Y.-C. N. Cheng and E. M. Haacke. *Magnetic Resonance Imaging*, volume 53. John Wiley & Sons Ltd, Chichester, UK, 2014.

[42] V. Cheplygina, L. Sørensen, and D. Tax. Classification of COPD with Multiple Instance Learning. *International conference on pattern recognition*, 2014.

[43] G. E. Christensen, J. H. Song, W. Lu, I. El Naqa, and D. A. Low. Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. *Medical physics*, 34(6):2155–63, 2007.

[44] H. O. Coxson, A. Dirksen, L. D. Edwards, J. C. Yates, A. Agusti, P. Bakke, P. M. Calverley, B. Celli, C. Crim, A. Duvoix, P. N. Fauerbach, D. A. Lomas, W. MacNee, R. J. Mayer, B. E. Miller, N. L. Müller, S. I. Rennard, E. K. Silverman, R. Tal-Singer, E. F. Wouters, and J. Vestbo. The presence and progression of emphysema in COPD as determined by CT scanning and biomarker expression: a prospective analysis from the ECLIPSE study. *The Lancet Respiratory Medicine*, 1(2):136–129, 2013.

[45] H. O. Coxson, J. Leipsic, G. Parraga, and D. D. Sin. Using pulmonary imaging to move chronic obstructive pulmonary disease beyond FEV1. *American journal of respiratory and critical care medicine*, 190(2):135–144, 2014.

[46] H. O. Coxson, J. Leipsic, G. Parraga, and D. D. Sin. Using pulmonary imaging to move chronic obstructive pulmonary disease beyond FEV1. *American journal of respiratory and critical care medicine*, 190(2):135–144, 2014.

[47] M. Deimling, V. Jellus, B. Geiger, and C. Chefd'hotel. Time Resolved Lung Ventilation Imaging by Fourier Decomposition. *Proceedings 16th Scientific Meeting, International Society for Magnetic Resonance in Medicine*, Toronto(c):2639, 2008.

[48] A. Depeursinge, A. Foncubierta-Rodriguez, D. Van de Ville, and H. Müller. Lung texture classification using locally-oriented Riesz components. *Medical image computing and computer-assisted intervention and Computer-Assisted Intervention (MICCAI)*, 14(Pt 3):231–8, 2011.

[49] A. A. Diaz, B. Bartholmai, R. San José Estépar, J. Ross, S. Matsuoka, T. Yamashiro, H. Hatabu, J. J. Reilly, E. K. Silverman, and G. R. Washko. Relationship of emphysema and airway disease assessed by CT to exercise capacity in COPD. *Respiratory Medicine*, 104(8):1145–1151, 2010.

[50] A. A. Diaz, C. E. Come, J. C. Ross, R. San José Estépar, M. K. Han, S. H. Loring, E. K. Silverman, and G. R. Washko. Association between airway caliber changes with lung inflation and emphysema assessed by volumetric CT scan in subjects with COPD. *Chest*, 141(3):736–744, 2012.

[51] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.

[52] T. Doel. Pulmonary Toolkit: code.google.com/p/pulmonarytoolkit/, 2017.

[53] T. Doel, D. J. Gavaghan, and V. Grau. Review of automatic pulmonary lobe segmentation methods from CT. *Computerized Medical Imaging and Graphics*, 40:13–29, 2015.

[54] T. Doel, T. N. Matin, F. V. Gleeson, D. J. Gavaghan, and V. Grau. Pulmonary lobe segmentation from CT images using fissureness, airways, vessels and multilevel B-splines. *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1491–1494, 2012.

[55] M. C. Donohue, H. Jacqmin-Gadda, M. Le Goff, R. G. Thomas, R. Raman, A. C. Gamst, L. A. Beckett, C. R. Jack, M. W. Weiner, J. F. Dartigues, and P. S. Aisen. Estimating long-term multivariate progression from short-term data, 2014.

[56] B. Driehuys, G. P. Cofer, J. Pollaro, J. B. Mackel, L. W. Hedlund, and G. A. Johnson. Imaging alveolar-capillary gas transfer using hyperpolarized 129Xe MRI. *Proceedings of the National Academy of Sciences*, 103(48):18278–18283, 2006.

[57] G. B. Drummond. Computed tomography and pulmonary measurements. *British journal of anaesthesia*, 80(5):665–671, 1998.

[58] M. B. Drummond, N. N. Hansel, J. E. Connett, P. D. Scanlon, D. P. Tashkin, and R. A. Wise. Spirometric predictors of lung function decline and mortality in early chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 185(12):1301–1306, 2012.

[59] S. Eda, K. Kubo, K. Fujimoto, Y. Matsuzawa, M. Sekiguchi, and F. Sakai. The relations between expiratory chest CT using helical CT and pulmonary function tests in emphysema. *American Journal of Respiratory and Critical Care Medicine*, 155(4):1290–1294, 1997.

[60] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, and H. Handels. Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. *IEEE transactions on medical imaging*, 30(2):251–65, 2011.

[61] R. S. J. Estépar, G. L. Kinney, J. L. Black-Shinn, R. P. Bowler, G. L. Kindlmann, J. C. Ross, R. Kikinis, M. K. Han, C. E. Come, A. A. Diaz, M. H. Cho, C. P. Hersh, J. D. Schroeder, J. J. Reilly, D. A. Lynch, J. D. Crapo, J. M. Wells, M. T. Dransfield, J. E. Hokanson, and G. R. Washko. Computed tomographic measures of pulmonary vascular morphology in smokers and their clinical implications. *American Journal of Respiratory and Critical Care Medicine*, 188(2):231–239, 2013.

[62] S. Fain, M. L. Schiebler, D. G. McCormack, and G. Parraga. Imaging of lung function using hyperpolarized helium-3 magnetic resonance imaging: Review of current and emerging translational methods and applications. *Journal of Magnetic Resonance Imaging*, 32(6):1398–1408, 2010.

[63] S. Fain, M. L. Schiebler, D. G. McCormack, and G. Parraga. Imaging of lung function using hyperpolarized helium-3 magnetic resonance imaging: Review of current and emerging translational methods and applications. *Journal of magnetic resonance imaging*, 32(6):1398–408, 2010.

[64] S. B. Fain, S. R. Panth, M. D. Evans, A. L. Wentland, J. H. Holmes, F. R. Korosec, M. J. O'Brien, H. Fountaine, and T. M. Grist. Early emphysematous changes in asymptomatic smokers: detection with 3He MR imaging. *Radiology*, 239(3):875–83, 2006.

[65] A. Feragen, J. Petersen, D. Grimm, A. Dirksen, J. H. Pedersen, K. Borgwardt, and M. de Bruijne. Geometric tree kernels: classification of COPD from airway tree geometry. *Information processing in medical imaging (IPMI)*, 23:171–83, 2013.

[66] A. Fishman, F. Martinez, and K. Naunheim. A Randomized Trial Comparing Lung-VolumeReduction Surgery with Medical Therapy for Severe Emphysema. *New England Journal of Medicine*, 348(21):2059–2073, 2003.

[67] C. Fletcher and R. Peto. The natural history of chronic airflow obstruction. *BMJ*, 1(6077):1645–1648, 1977.

[68] H. M. Fonteijn, M. Modat, M. J. Clarkson, J. Barnes, M. Lehmann, N. Z. Hobbs, R. I. Scahill, S. J. Tabrizi, S. Ourselin, N. C. Fox, and D. C. Alexander. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage*, 60(3):1880–1889, 2012.

[69] Z. G. Fridlender, R. Kaufman, D. Shaham, and N. Berkman. Can emphysema predict exacerbations and changes in PFTs in COPD? *European Respiratory Journal*, 48(suppl 60), 2016.

[70] C. J. Galbán, M. K. Han, J. L. Boes, K. A. Chughtai, R. Charles, T. D. Johnson, S. Galbán, A. Rehemtulla, E. A. Kazerooni, F. J. Martinez, and B. D. Ross. CT-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nature Medicine*, 18(11):1711–1715, 2013.

[71] C. J. Galbán, M. K. Han, J. L. Boes, K. a. Chughtai, C. R. Meyer, T. D. Johnson, S. Galbán, A. Rehemtulla, E. a. Kazerooni, F. J. Martinez, and B. D. Ross. Com-

puted tomographybased biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nature Medicine*, 18(11):1711–1716, 2012.

[72] L. Gallardo-Estrella, D. A. Lynch, M. Prokop, D. Stinson, J. Zach, P. F. Judy, B. van Ginneken, and E. M. van Rikxoort. Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification. *European Radiology*, 26(2):478–486, 2016.

[73] J. Garcia-Aymerich, F. P. Gómez, M. Benet, E. Farrero, X. Basagaña, A. Gayete, C. Paré, X. Freixa, J. Ferrer, A. Ferrer, J. Roca, J. B. Gáldiz, J. Sauleda, E. Monsó, J. Gea, J. a. Barberà, A. Agustí, and J. M. Antó. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. *Thorax*, 66:430–437, 2010.

[74] P. A. Gevenois, P. De Vuyst, V. De Maertelaer, J. Zanen, D. Jacobovitz, M. G. Cosio, and J. C. Yernault. Comparison of computed density and microscopic morphometry in pulmonary emphysema. *American Journal of Respiratory and Critical Care Medicine*, 154(1):187–192, 1996.

[75] H. A. Gietema, P. Zanen, A. Schilham, B. van Ginneken, R. J. van Klaveren, M. Prokop, and J. W. J. Lammers. Distribution of emphysema in heavy smokers: Impact on pulmonary function. *Respiratory Medicine*, 104(1):76–82, 2010.

[76] G. González, S. Y. Ash, G. Vegas Sanchez-Ferrero, J. Onieva Onieva, F. N. Rahaghi, J. C. Ross, A. Díaz, R. San José Estépar, G. R. Washko, f. t. C. Investigators, and ECLIPSE. Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography. *American Journal of Respiratory and Critical Care Medicine*, 197(2):193–203, 2017.

[77] V. Gorbunova, S. S. A. M. Jacobs, P. Lo, A. Dirksen, M. Nielsen, A. Bab-Hadiashar, and M. de Bruijne. Early detection of emphysema progression. In *Medical image computing and computer-assisted intervention*, volume 13, pages 193–200, 2010.

[78] T. B. Grydeland, A. Dirksen, H. O. Coxson, S. G. Pillai, S. Sharma, G. E. Eide, A. Gulsvik, and P. S. Bakke. Quantitative CT: Emphysema and airway wall thickness by gender, age and smoking. *European Respiratory Journal*, 34(4), 2009.

[79] T. B. Grydeland, E. Thorsen, A. Dirksen, R. Jensen, H. O. Coxson, S. G. Pillai, S. Sharma, G. E. Eide, A. Gulsvik, and P. S. Bakke. Quantitative CT measures of emphysema and airway wall thickness are related to DLCO. *Respiratory Medicine*, 105(3):343–351, 2011.

[80] S. Gu, C. Fuhrman, X. Meng, J. M. Siegfried, D. Gur, J. K. Leader, F. C. Sciurba, and J. Pu. Computerized identification of airway wall in CT examinations using a 3D active surface evolution approach. *Medical Image Analysis*, 17(3):283–296, 2013.

[81] J. W. Gurney, K. K. Jones, R. A. Robbins, G. L. Gossman, K. J. Nelson, D. Daughton, J. R. Spurzem, and S. I. Rennard. Regional distribution of emphysema: correlation of high-resolution CT with pulmonary function tests in unselected smokers. *Radiology*, 183(2):457–63, 1992.

[82] J. W. Gurney, K. K. Jones, R. A. Robbins, G. L. Gossman, K. J. Nelson, D. Daughton, J. R. Spurzem, and S. I. Rennard. Regional distribution of emphysema: correlation of high-resolution CT with pulmonary function tests in unselected smokers. *Radiology*, 183(2):457–63, 1992.

[83] M. Hackx, A. a. Bankier, and P. A. Gevenois. Chronic obstructive pulmonary disease: CT quantification of airways disease. *Radiology*, 265(1):34–48, 2012.

[84] Y. Hame, E. D. Angelini, E. A. Hoffman, R. G. Barr, and A. F. Laine. Adaptive quantification and longitudinal analysis of pulmonary emphysema with a hidden Markov measure field model. *IEEE transactions on medical imaging*, 33(7):1527–40, 2014.

[85] M. K. Han, A. Agusti, P. M. Calverley, B. R. Celli, G. Criner, J. L. Curtis, L. M. Fabbri, J. G. Goldin, P. W. Jones, W. Macnee, B. J. Make, K. F. Rabe, S. I. Rennard, F. C. Sciurba, E. K. Silverman, J. Vestbo, G. R. Washko, E. F. M. Wouters, and F. J. Martinez. Chronic obstructive pulmonary disease phenotypes: the future of COPD. *American journal of respiratory and critical care medicine*, 182(5):598–604, 2010.

[86] M. K. Han, A. Agusti, P. M. Calverley, B. R. Celli, G. Criner, J. L. Curtis, L. M. Fabbri, J. G. Goldin, P. W. Jones, W. Macnee, B. J. Make, K. F. Rabe, S. I. Rennard, F. C. Sciurba, E. K. Silverman, J. Vestbo, G. R. Washko, E. F. M. Wouters, and F. J. Martinez. Chronic obstructive pulmonary disease phenotypes: the future of COPD. *American journal of respiratory and critical care medicine*, 182(5):598–604, 2010.

[87] M. K. Han, E. A. Kazerooni, D. A. Lynch, L. X. Liu, S. Murray, J. L. Curtis, G. J. Criner, V. Kim, R. P. Bowler, N. A. Hanania, A. R. Anzueto, B. J. Make, J. E. Hokanson, J. D. Crapo, E. K. Silverman, F. J. Martinez, and G. R. Washko. Chronic obstructive pulmonary disease exacerbations in the COPDGene study: associated radiologic phenotypes. *Radiology*, 261(1):274–82, 2011.

[88] R. Harmouche, J. C. Ross, A. A. Diaz, G. R. Washko, and R. S. J. Estepar. A Robust Emphysema Severity Measure Based on Disease Subtypes. *Academic Radiology*, 23(4):421–428, 2016.

[89] R. Harmouche, J. C. Ross, G. R. Washko, and R. S. J. Estepar. Pectoralis muscle segmentation on ct images based on Bayesian graph cuts with a subject-tailored atlas. In *Medical computer vision: algorithms for big data (MCV)*, volume 8848, pages 34–44. Springer, Cham, 2014.

[90] H. Hayashi, K. Nagaoki, and H. Otsuji. Radiographic and CT Major Fissures 1. *Radiographics*, pages 861–874, 2001.

[91] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel. MIND: Modality Independent Neighbourhood Descriptor for Multi-Modal Deformable Registration. *Medical Image Analysis*, 16(7):1423–1435, 2012.

[92] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, S. M. Brady, and J. A. Schnabel. MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical image analysis*, 16(7):1423–35, 2012.

[93] M. P. Heinrich, M. Jenkinson, M. Brady, and J. A. Schnabel. MRF-Based Deformable Registration and Ventilation Estimation of Lung CT. *IEEE transactions on medical imaging*, 32(7):1239–48, jul 2013.

[94] C. P. Hersh, G. R. Washko, F. L. Jacobson, R. Gill, R. San Jose Estepar, J. J. Reilly, and E. K. Silverman. Interobserver variability in the determination of upper lobe-predominant emphysema. *Chest*, 131(2):424–431, 2007.

[95] E. A. Hoffman. Effect of body orientation on regional lung expansion: a computed tomographic approach. *Journal of applied physiology*, 59(2):468–480, 1985.

[96] J. Hogg, F. Chu, and S. Utokaparch. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *New England journal of medicine*, 350(26):2645–2653, 2004.

[97] J. C. Hogg, P. T. Macklem, and W. M. Thurlbeck. Site and nature of airway obstruction in chronic obstructive lung disease. *The New England journal of medicine*, 278(25):1355–60, 1968.

[98] J. C. Hogg, J. E. McDonough, and M. Suzuki. Small airway obstruction in COPD: new insights based on micro-CT imaging and MRI imaging. *Chest*, 143(5):1436–1443, 2013.

[99] Https://grand-challenge.org/site/lola11/. LOLA11.

[100] S. Hu, E. A. Hoffman, and J. M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE transactions on medical imaging*, 20(6):490–8, 2001.

[101] J. R. Hurst, J. Vestbo, A. Anzueto, N. Locantore, H. Müllerova, R. Tal-Singer, B. Miller, D. A. Lomas, A. Agusti, W. Macnee, P. Calverley, S. Rennard, E. F. M. Wouters, and J. A. Wedzicha. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *The New England journal of medicine*, 363(12):1128–38, 2010.

[102] B. M. Jedynak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jedynak, B. Caffo, and J. L. Prince. A computational neurodegenerative disease progression score: Method and results with the Alzheimer's disease neuroimaging initiative cohort. *NeuroImage*, 63(3):1478–1486, 2012.

[103] K. M. Johnson, S. B. Fain, M. L. Schiebler, and S. Nagle. Optimized 3D ultrashort echo time pulmonary MRI. *Magnetic Resonance in Medicine*, 70(5):1241–1250, 2013.

[104] J. Ju, R. Li, S. Gu, J. K. Leader, X. Wang, Y. Chen, B. Zheng, S. Wu, D. Gur, F. Sciurba, and J. Pu. Impact of emphysema heterogeneity on pulmonary function. *PLoS ONE*, 9(11), 2014.

[105] M. Kirby, N. Kanhere, R. Etemad-Rezai, D. G. McCormack, and G. Parraga. Hyperpolarized Helium-3 magnetic resonance imaging of chronic obstructive pulmonary disease exacerbation. *Journal of magnetic resonance imaging*, 37(5):1223–7, 2013.

[106] M. Kirby, S. Svenningsen, N. Kanhere, A. Owrangi, A. Wheatley, H. O. Coxson, G. E. Santyr, N. a. M. Paterson, D. G. McCormack, and G. Parraga. Pulmonary ventilation visualized using hyperpolarized helium-3 and xenon-129 magnetic resonance imaging: differences in COPD and relationship to emphysema. *Journal of applied physiology*, 114(6):707–15, 2013.

[107] S. Kiryu, T. Sundaram, S. Kubo, K. Ohtomo, T. Asakura, J. C. Gee, H. Hatabu, and M. Takahashi. MRI assessment of lung parenchymal motion in normal mice and transgenic mice with sickle cell disease. *Journal of magnetic resonance imaging*, 27(1):49–56, 2008.

[108] T. Klinder, H. Wendland, and R. Wiemker. Lobar fissure detection using line enhancing filters. *SPIE Medical imaging*, 8669:86693C, 2013.

[109] B. Lassen and E. van Rikxoort. Automatic segmentation of the pulmonary lobes from chest CT scans based on fissures, vessels, and bronchi. *IEEE transactions on medical imaging*, 32(2):210–22, 2013.

[110] E. Levina and P. Bickel. The earth mover's distance is the Mallows distance: some insights from statistics. *Eighth IEEE International Conference on Computer Vision*, 2:251–256, 2001.

[111] J. Ley-Zaporozhan, S. Ley, R. Eberhardt, O. Weinheimer, C. Fink, M. Puderbach, M. Eichinger, F. Herth, and H. U. Kauczor. Assessment of the relationship between lung parenchymal destruction and impaired pulmonary perfusion on a lobar level in patients with emphysema. *European Journal of Radiology*, 63(1):76–83, 2007.

[112] B. Li, G. E. Christensen, E. A. Hoffman, G. McLennan, and J. M. Reinhardt. Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images. *Academic radiology*, 10(3):255–65, 2003.

[113] B. Li, G. E. Christensen, E. A. Hoffman, G. McLennan, and J. M. Reinhardt. Establishing a normative atlas of the human lung: computing the average transformation and atlas construction. *Academic radiology*, 19(11):1368–81, 2012.

[114] C. Liguori, G. Frauenfelder, C. Massaroni, P. Saccomandi, F. Giurazza, F. Pitocco, R. Marano, and E. Schena. Emerging clinical applications of computed tomography. *Medical devices: evidence and research*, 8:265–78, 2015.

[115] C.-L. Lin, M. H. Tawhai, G. McLennan, and E. a. Hoffman. Multiscale simulation of gas flow in Subject-Specific Models of the Human Lung. *Wiley interdisciplinary reviews: systems biology and medicine*, 5(5):643–655, 2013.

[116] X. Liu, I. Oguz, S. M. Pizer, and G. S. Mageras. Shape-correlated deformation statistics for respiratory motion prediction in 4D lung. *Proceedings of SPIE*, 7625:76252D–76252D–10, 2010.

[117] P. Lo, B. van Ginneken, J. Reinhardt, and M. de Bruijne. Extraction of Airways From CT (EXACT'09). *IEEE transactions on medical imaging*, 31(11):2093–2107, jul 2012.

[118] P. Lo, E. M. van Rikxoort, F. Abtin, S. Ahmad, A. Ordookhani, J. Goldin, and M. S. Brown. Automated segmentation of pulmonary lobes in chest CT scans using evolving surfaces. 8669, 2013.

[119] O. Lucidarme, E. Coche, P. Cluzel, I. Mourey-Gerosa, N. Howarth, and P. Grenier. Expiratory CT scans for chronic airway disease: Correlation with pulmonary function test results. *American Journal of Roentgenology*, 170(2):301–307, 1998.

[120] G. Lutterbey, J. Gieseke, M. von Falkenhausen, N. Morakkabati, and H. Schild. Lung MRI at 3.0 T: A comparison of helical CT and high-field MRI in the detection of diffuse lung disease. *European Radiology*, 15(2):324–328, 2005.

[121] D. A. Lynch and M. A. Al-Qaisi. Quantitative Computed Tomography in Chronic Obstructive Pulmonary Disease. *Journal of Thoracic Imaging*, 28(5):284–290, 2013.

[122] W. Ma, K. Sheikh, S. Svenningsen, D. Pike, F. Guo, R. Etemad-Rezai, J. Leipsic, H. O. Coxson, D. G. McCormack, and G. Parraga. Ultra-short echo-time pulmonary MRI: Evaluation and reproducibility in COPD subjects with and without bronchiectasis. *Journal of Magnetic Resonance Imaging*, 41(5):1465–1474, 2015.

[123] P. T. Macklem. The physiology of small airways. *American journal of respiratory and critical care medicine*, 157(5 Pt 2):S181–3, 1998.

[124] W. MacNee and P. M. Calverley. Chronic obstructive pulmonary disease . 7: Management of COPD. *Thorax*, 58(3):261–265, 2003.

[125] C. H. Martinez, Y.-H. Chen, P. M. Westgate, L. X. Liu, S. Murray, J. L. Curtis, B. J. Make, E. A. Kazerooni, D. A. Lynch, N. Marchetti, G. R. Washko, F. J. Martinez, and M. K. Han. Relationship between quantitative CT metrics and health status and BODE in chronic obstructive pulmonary disease. *Thorax*, 67(5):399–406, 2012.

[126] S. Matsuoka, Y. Kurihara, K. Yagihashi, M. Hoshino, N. Watanabe, and Y. Nakajima. Quantitative assessment of air trapping in chronic obstructive pulmonary disease using inspiratory and expiratory volumetric MDCT. *American journal of roentgenology*, 190(3):762–9, 2008.

[127] J. R. Mayo, A. MacKay, and N. L. Muller. MR imaging of the lungs: Value of short TE spin-echo pulse sequences. *American Journal of Roentgenology*, 159(5):951–956, 1992.

[128] J. R. McClelland, D. J. Hawkes, T. Schaeffter, and A. P. King. Respiratory motion models: A review. *Medical image analysis*, 17(1):42–19, 2012.

[129] J. R. Mcclelland, M. Modat, S. Arridge, H. Grimes, D. D. Souza, D. Thomas, D. O. Connell, D. A. Low, E. Kaza, D. J. Collins, M. O. Leach, and D. J. Hawkes. A generalized framework unifying image registration and respiratory motion models and incorporating image reconstruction, for partial or full imaging data. *Physics in medicine and biology*, 2017.

[130] A. Mechelli, C. J. Price, K. J. Friston, and J. Ashburner. Voxel-based morphometry of the human brain: methods and applications. *Current Medical Imaging Reviews*, 1(2):105–113, 2005.

[131] C. S. Mendoza, G. R. Washko, J. C. Ross, A. A. Diaz, D. A. Lynch, and J. D. Crapo. Emphysema quantification in a multi-scanner HRCT cohort using local intensity distributions. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 474–477, 2012.

[132] O. M. Mets, P. A. de Jong, B. van Ginneken, H. A. Gietema, and J. W. J. Lammers. Quantitative computed tomography in COPD: possibilities and limitations. *Lung*, 190(2):133–45, 2012.

[133] M. Miravitlles, M. Calle, and J. J. Soler-Cataluna. Clinical Phenotypes of COPD: Identification, Definition and Implications for Guidelines. *Archivos de Bronconeumologia*, 48(3):86–98, 2012.

[134] M. Mishima, T. Hirai, H. Itoh, Y. Nakano, H. Sakai, S. Muro, K. Nishimura, Y. Oku, K. Chin, M. Ohi, T. Nakamura, J. H. Bates, A. M. Alencar, and B. Suki. Complexity of terminal airspace geometry assessed by lung computed tomography in normal subjects and patients with chronic obstructive pulmonary disease. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16):8829–34, 1999.

[135] M. Modat, D. M. Cash, P. Daga, G. P. Winston, J. S. Duncan, and S. Ourselin. Global image registration using a symmetric block-matching approach. *Journal of Medical Imaging*, 1(2):024003, 2014.

[136] M. Modat, J. McClelland, and S. Ourselin. Lung registration using the NiftyReg package. *Medical Image Analysis for the Clinic: A Grand Challenge EMPIRE 10*, pages 33–42, 2010.

[137] M. Modat, J. McClelland, and S. Ourselin. Lung Registration Using the NiftyReg Package. *Medical Image Analysis for the Clinic: A Grand Challenge EMPIRE 10*, pages 33–42, 2010.

[138] M. Modat, G. R. Ridgway, P. Daga, M. J. Cardoso, D. J. Hawkes, J. Ashburner, and S. Ourselin. Log-Euclidean free-form deformation. In B. M. Dawant and D. R. Haynor, editors, *SPIE*, volume 7962, 2011.

[139] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3):278–84, 2010.

[140] K. Mori, J. Hasegawa, J. Toriwaki, H. Anno, and K. Katada. Recognition of bronchus in three-dimensional X-ray CT images with applications to virtualized bronchoscopy system. In *Proceedings - International Conference on Pattern Recognition*, volume 3, pages 528–532. IEEE, 1996.

[141] J. P. Mugler and T. A. Altes. Hyperpolarized 129Xe MRI of the human lung. *Journal of Magnetic Resonance Imaging*, 37(2):313–331, 2013.

[142] N. L. Muller, C. A. Staples, R. R. Miller, and R. T. Abboud. 'Density mask'. An objective method to quantitate emphysema using computed tomography. *Chest*, 94(4):782–787, 1988.

[143] K. Murphy, J. P. W. Pluim, E. M. van Rikxoort, P. A. de Jong, B. de Hoop, H. A. Gietema, O. Mets, M. de Bruijne, P. Lo, M. Prokop, and B. van Ginneken. Toward automatic regional analysis of pulmonary function using inspiration and expiration thoracic CT. *Medical physics*, 39(3):1650–62, 2012.

[144] K. Murphy, B. van Ginneken, and S. Klein et al. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE transactions on medical imaging*, 30(11):1901–1920, 2011.

[145] E. A. Nadaraya. On Estimating Regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

[146] D. P. Naidich. *Imaging of the Airways: functional and radiologic correlations*. 2005.

[147] Y. Nakano, S. Muro, H. Sakai, T. Hirai, K. Chin, M. Tsukino, K. Nishimura, H. Itoh, P. D. Paré, J. C. Hogg, and M. Mishima. Computed tomographic measurements of airway dimensions and emphysema in smokers. Correlation with lung function. *American journal of respiratory and critical care medicine*, 162(3 Pt 1):1102–8, 2000.

[148] Y. Nakano, J. C. Wong, P. A. De Jong, L. Buzatu, T. Nagao, H. O. Coxson, W. M. Elliott, J. C. Hogg, and P. D. Paré. The prediction of small airway dimensions using computed tomography. *American Journal of Respiratory and Critical Care Medicine*, 171(2):142–146, 2005.

[149] K. B. Newman, D. A. Lynch, L. S. Newman, D. Ellegood, and J. D. Newell. Quantitative computed tomography detects air trapping due to asthma. *Chest*, 106(1):105–109, 1994.

[150] Y. Nimura, T. Kitasaka, H. Honma, H. Takabatake, M. Mori, H. Natori, and K. Mori. Lung lobe segmentation based on statistical atlas and graph cuts. *SPIE Proceedings*, 8315, 2012.

[151] R. A. Ochs, J. G. Goldin, F. Abtin, H. J. Kim, K. Brown, P. Batra, D. Roback, M. F. McNitt-Gray, and M. S. Brown. Automated classification of lung bronchovascular anatomy in CT using AdaBoost. *Medical Image Analysis*, 11(3):315–324, 2007.

[152] Y. Ohno, H. Koyama, T. Yoshikawa, K. Matsumoto, M. Takahashi, M. Van Cauteren, and K. Sugimura. T2* measurements of 3-T MRI with ultrashort TEs: Capabilities of pulmonary function assessment and clinical stage classification in smokers. *American Journal of Roentgenology*, 197(2):W279–W285, 2011.

[153] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[154] S. Ourselin, A. Roche, S. Prima, and N. Ayache. Block Matching : A General Framework to Improve Robustness of Rigid Registration of Medical Images. *Medical image computing and computer-assisted intervention (MICCAI)*, pages 557–566, 2000.

[155] B. W. Papiez, M. P. Heinrich, J. Fehrenbach, L. Risser, and J. A. Schnabel. An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. *Medical Image Analysis*, 18(8):1299–1311, 2014.

[156] D. G. Parr, B. Stoel, J. Stolk, and R. Stockley. Validation of computed tomographic lung densitometry for monitoring emphysema in alpha1-antitrypsin deficiency. *Thorax*, 61(6):485–490, 2006.

[157] D. G. Parr, B. C. Stoel, J. Stolk, and R. a. Stockley. Pattern of emphysema distribution in alpha1-antitrypsin deficiency influences lung function impairment. *American journal of respiratory and critical care medicine*, 170(11):1172–8, 2004.

[158] A. R. C. Patel and J. R. Hurst. Extrapulmonary comorbidities in chronic obstructive pulmonary disease: state of the art. *Expert review of respiratory medicine*, 5(5):647–62, 2011.

[159] J. H. Pedersen, H. Ashraf, A. Dirksen, K. Bach, H. Hansen, P. Toennesen, H. Thorsen, J. Brodersen, B. G. Skov, M. DOøssing, J. Mortensen, K. Richter, P. Clementsen, and N. Seersholm. The danish randomized lung cancer ct screening trial- overall design and results of the prevalence round. *Journal of Thoracic Oncology*, 4(5):608–614, 2009.

[160] O. Pele and M. Werman. The quadratic-chi histogram distance family. In *Lecture Notes in Computer Science*, volume 6312 LNCS, pages 749–762, 2010.

[161] J. Petersen, M. M. W. Wille, L. L. Rakêt, A. Feragen, J. H. Pedersen, M. Nielsen, A. Dirksen, and M. de Bruijne. Effect of inspiration on airway dimensions measured in maximal inspiration CT images of subjects without airflow limitation. *European radiology*, 24(9):2319–2325, 2014.

[162] D. Pike, M. Kirby, R. L. Eddy, F. Guo, D. P. I. Capaldi, A. Ouriadov, D. G. McCormack, and G. Parraga. Regional Heterogeneity of Chronic Obstructive Pulmonary Disease Phenotypes: Pulmonary (3)He Magnetic Resonance Imaging and Computed Tomography. *COPD*, 13(5):601–609, 2016.

[163] E. Pompe, E. M. Van Rikxoort, M. Schmidt, J. Ruhaak, L. Gallardo Estrella, R. Vliegenthart, M. Oudkerk, H. J. De Koning, B. Van Ginneken, P. A. De Jong, J. W. J. Lammers, and F. A. A. Mohamed Hoesein. Parametric response mapping adds value to current computed tomography biomarkers in diagnosing chronic obstructive pulmonary disease, 2015.

[164] J. Pu, B. Zheng, J. K. Leader, C. Fuhrman, F. Knollmann, A. Klym, and D. Gur. Pulmonary lobe segmentation in CT examinations using implicit surface fitting. *IEEE Transactions on Medical Imaging*, 28(12):1986–1996, 2009.

[165] K. F. Rabe, S. Hurd, A. Anzueto, P. J. Barnes, S. A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. van Weel, and J. Zielinski. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *American journal of respiratory and critical care medicine*, 176(6):532–55, 2007.

[166] E. A. Regan, J. E. Hokanson, J. R. Murphy, B. Make, D. A. Lynch, T. H. Beaty, D. Curran-Everett, E. K. Silverman, and J. D. Crapo. Genetic epidemiology of COPD (COPDGene) study design. *COPD*, 7(1):32–43, 2010.

[167] E. A. Regan, J. E. Hokanson, J. R. Murphy, B. Make, D. A. Lynch, T. H. Beaty, D. Curran-Everett, E. K. Silverman, and J. D. Crapo. Genetic epidemiology of COPD (COPDGene) study design. *COPD*, 7(1):32–43, 2010.

[168] J. M. Reinhardt, K. Ding, G. E. Christensen, E. A. Hoffman, and S. V. Bodas. Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation. *Medical image analysis*, 12(6):752–63, 2008.

[169] J. M. Reinhardt, N. D. D'Souza, and E. a. Hoffman. Accurate measurement of intrathoracic airways. *IEEE transactions on medical imaging*, 16(6):820–827, 1997.

[170] S. I. Rennard and S. G. Farmer. Exacerbations and progression of disease in asthma and chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, 1(2):88–92, 2004.

[171] L. Risser, F. X. Vialard, H. Y. Baluwala, and J. A. Schnabel. Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions. *Medical Image Analysis*, 17(2):182–193, 2013.

[172] M. a. V. Ro F. Frangi, Wiro J. Niessen, Koen L. Vincken. Multiscale vessel enhancement filtering. *Medical image computing and computer-assisted intervention (MICCAI)*, 1496:130–137, 1998.

[173] R. Rodríguez-Roisin. COPD exacerbations 5: Management. *Thorax*, 61(6):535–44, 2006.

[174] J. C. Ross, G. L. Kindlmann, Y. Okajima, H. Hatabu, A. a. Díaz, E. K. Silverman, G. R. Washko, J. Dy, and R. San José Estépar. Pulmonary lobe segmentation based on ridge surface sampling and shape model fitting. *Medical physics*, 40(12):121903, 2013.

[175] Y. Rubner, C. Tomasi, and L. J. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[176] R. D. Rudyanto, S. Kerkstra, E. M. van Rikxoort, C. Fetita, P. Y. Brillet, C. Lefevre, W. Xue, X. Zhu, J. Liang, I. lkay Oksuz, D. Unay, K. Kadipasaoglu, R. S. J. Estepar, J. C. Ross, G. R. Washko, J. C. Prieto, M. H. Hoyos, M. Orkisz, H. Meine, M. H??llebrand, C. Stocker, F. L. Mir, V. Naranjo, E. Villanueva, M. Staring, C. Xiao, B. C. Stoel, A. Fabijanska, E. Smistad, A. C. Elster, F. Lindseth, A. H. Foruzan, R. Kiros, K. Popuri, D. Cobzas, D. Jimenez-Carretero, A. Santos, M. J. Ledesma-Carbayo, M. Helmberger, M. Urschler, M. Pienn, D. G. H. Bosboom, A. Campo,

M. Prokop, P. A. de Jong, C. Ortiz-de Solorzano, A. Munoz-Barrutia, and B. van Ginneken. Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: The VESSEL12 study. *Medical Image Analysis*, 18(7):1217–1232, 2014.

[177] J. Ruhaak, T. Polzin, S. Heldmann, I. Simpson, H. Handels, J. Modersitzki, and M. P. Heinrich. Estimation of Large Motion in Lung CT by Integrating Regularized Keypoint Correspondences into Dense Deformable Registration. *IEEE Transactions on Medical Imaging*, pages 1–1, 2017.

[178] E. P. A. Rutten, T. B. Grydeland, S. G. Pillai, S. Wagers, A. Dirksen, H. O. Coxson, A. Gulsvik, E. F. M. Wouters, and P. S. Bakke. Quantitative CT: Associations between emphysema, airway wall thickness and body composition in COPD. *Pulmonary Medicine*, 2011:1–6, 2011.

[179] B. T. Saam. Magnetic resonance imaging with laser-polarized noble gases. *Nature Medicine*, 2(3):358–359, 1996.

[180] T. Saitoh, H. Koba, N. Shijubo, H. Tanaka, and F. Sugaya. Lobar distribution of emphysema in computed tomographic densitometric analysis. *Investigative Radiology*, 35(4):235–243, 2000.

[181] K. Sashidhar, M. Gulati, D. Gupta, S. Monga, and S. Suri. Emphysema in heavy smokers with normal chest radiography. Detection and quantification by HCRT. *Acta radiologica*, 43(1):60–5, 2002.

[182] A. Schmidt-Richberg, J. Ehrhardt, R. Werner, and H. Handels. Lung registration with improved fissure alignment by integration of pulmonary lobe segmentation. *Medical image computing and computer-assisted intervention (MICCAI)*, 15(Pt 2):74–81, 2012.

[183] A. Schmidt-Richberg, C. Ledig, R. Guerrero, H. Molina-Abril, A. Frangi, and D. Rueckert. Learning biomarker models for progression estimation of Alzheimer's disease. *PLoS ONE*, 11(4):e0153040, 2016.

[184] A. Schmidt-Richberg, R. Werner, H. Handels, and J. Ehrhardt. Estimation of slipping organ motion by registration with direction-dependent regularization. *Medical image analysis*, 16(1):150–9, 2012.

[185] J. D. Schroeder, A. S. McKenzie, J. a. Zach, C. G. Wilson, D. Curran-Everett, D. S. Stinson, J. D. Newell, and D. a. Lynch. Relationships between airflow obstruction and quantitative CT measurements of emphysema, air trapping, and airways in subjects with and without chronic obstructive pulmonary disease. *AJR. American journal of roentgenology*, 201(3):460–70, 2013.

[186] T. A. Seemungal, G. C. Donaldson, A. Bhowmik, D. J. Jeffries, and J. a. Wedzicha. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine*, 161(5):1608–13, 2000.

[187] P. L. Shah and F. J. F. Herth. Current status of bronchoscopic lung volume reduction with endobronchial valves. *Thorax*, 69(3):280–6, 2014.

[188] S. Siddiqui and O. S. Usmani. Small airways, big challenge: measuring the unseen? *Nature Medicine*, 18(11):1619–1621, 2012.

[189] I. Sluimer, A. Schilham, A. Prokop, and B. van Ginneken. Computer analysis of computed tomography scans of the lung: a survey. *IEEE transactions on medical imaging*, 25(4):385–405, 2006.

[190] N. Smith and A. Webb. *Introduction to medical imaging: physics, engineering, and clinical applications*. 2011.

[191] G. Song, N. Tustison, B. Avants, and J. C. Gee. Lung CT image registration using diffeomorphic transformation models. *Medical Image Analysis for the Clinic: A Grand Challenge EMPIRE 10*, 2010.

[192] L. Sørensen, M. Loog, P. Lo, H. Ashraf, A. Dirksen, R. P. W. Duin, and M. de Bruijne. Image dissimilarity-based quantification of lung disease from CT. In *Medical image computing and computer-assisted intervention (MICCAI)*, volume 13, pages 37–44, 2010.

[193] L. Sørensen, M. Nielsen, P. Lo, H. Ashraf, J. H. Pedersen, and M. de Bruijne. Texture-based analysis of COPD: a data-driven approach. *IEEE transactions on medical imaging*, 31(1):70–8, 2012.

[194] L. Sorensen, S. B. Shaker, and M. de Bruijne. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Transactions on Medical Imaging*, 29(2):559–569, 2010.

[195] J. Srinakarin, J. Thammaroj, and W. Boonsawat. Comparison of high-resolution computed tomography with pulmonary function testing in symptomatic smokers. *Journal of the Medical Association of Thailand*, 86(6):522–528, 2003.

[196] A. Stadler, P. M. Jakob, M. Griswold, L. Stiebellehner, M. Barth, and A. A. Bankier. T1 mapping of the entire lung parenchyma: Influence of respiratory phase and correlation to lung function test results in patients with diffuse lung disease. *Magnetic Resonance in Medicine*, 59(1):96–101, 2008.

[197] G. J. Stanisz, E. E. Odrobina, J. Pun, M. Escaravage, S. J. Graham, M. J. Bronskill, and R. M. Henkelman. T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magnetic Resonance in Medicine*, 54(3):507–512, 2005.

[198] M. Staring, M. E. Bakker, D. P. Shamonin, J. Stolk, J. H. C. Reiber, and B. C. Stoel. Towards local estimation of emphysema progression using image registration. In J. P. W. Pluim and B. M. Dawant, editors, *SPIE Medical Imaging*, pages 72590O–72590O–9. International Society for Optics and Photonics, 2009.

[199] B. C. Stoel, H. Putter, M. E. Bakker, A. Dirksen, R. A. Stockley, E. Piitulainen, E. W. Russi, D. Parr, S. B. Shaker, J. H. C. Reiber, and J. Stolk. Volume Correction in Computed Tomography Densitometry for Follow-up Studies on Pulmonary Emphysema. *Proceedings of the American Thoracic Society*, 5(9):919–924, 2008.

[200] D. R. Subramanian, S. Gupta, D. Burggraf, S. J. Vom Silberberg, I. Heimbeck, M. S. Heiss-Neumann, K. Haeussinger, C. Newby, B. Hargadon, V. Raj, D. Singh, U. Kolsum, T. P. Hofer, K. Al-Shair, N. Luetzen, A. Prasse, J. Müller-Quernheim, G. Benea, S. Leprotti, P. Boschetto, D. Gorecka, A. Nowinski, K. Oniszh, W. Zu Castell, M. Hagen, I. Barta, B. Döme, J. Strausz, T. Greulich, C. Vogelmeier, A. R. Koczulla, I. Gut, J. Hohlfeld, T. Welte, M. Lavae-Mokhtari, L. Ziegler-Heitbrock, C. Brightling, and D. G. Parr. Emphysema- and airway-dominant COPD phenotypes defined by standardised quantitative computed tomography. *European Respiratory Journal*, 48(1):92–103, 2016.

[201] B. Suki and J. H. T. Bates. Extracellular matrix mechanics in lung parenchymal diseases. *Respiratory physiology & neurobiology*, 163(1-3):33–43, 2008.

[202] B. Suki, K. R. Lutchen, and E. P. Ingenito. On the progressive nature of emphysema: roles of proteases, inflammation, and mechanical forces. *American journal of respiratory and critical care medicine*, 168(5):516–21, 2003.

[203] N. Tanabe, S. Muro, T. Hirai, T. Oguma, K. Terada, S. Marumo, D. Kinose, E. Ogawa, Y. Hoshino, and M. Mishima. Impact of exacerbations on emphysema progression in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine*, 183(12):1653–9, 2011.

[204] N. Tanabe, S. Muro, S. Sato, S. Tanaka, T. Oguma, H. Kiyokawa, T. Takahashi, D. Kinose, Y. Hoshino, T. Kubo, T. Hirai, and M. Mishima. Longitudinal study of spatially heterogeneous emphysema progression in current smokers with chronic obstructive pulmonary disease. *PloS one*, 7(9):e44993, 2012.

[205] N. Tanabe, S. Muro, S. Sato, S. Tanaka, T. Oguma, H. Kiyokawa, T. Takahashi, D. Kinose, Y. Hoshino, T. Kubo, T. Hirai, and M. Mishima. Longitudinal study of spatially heterogeneous emphysema progression in current smokers with chronic obstructive pulmonary disease. *PloS one*, 7(9):e44993, 2012.

[206] N. Tanabe, S. Muro, S. Tanaka, S. Sato, T. Oguma, H. Kiyokawa, T. Takahashi, D. Kinose, Y. Hoshino, T. Kubo, E. Ogawa, T. Hirai, and M. Mishima. Emphysema distribution and annual changes in pulmonary function in male patients with chronic obstructive pulmonary disease. *Respiratory research*, 13:31, 2012.

[207] C. Tantucci and D. Modina. Lung function decline in COPD. *International Journal of Chronic Obstructive Pulmonary Disease*, 7:95, 2012.

[208] E. D. Telenga, M. Oudkerk, P. M. A. Van Ooijen, R. Vliegenthart, N. H. T. Ten Hacken, D. S. Postma, and M. Van Den Berge. Airway wall thickness on HRCT scans decreases with age and increases with smoking. *BMC Pulmonary Medicine*, 17(27), 2017.

[209] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–23, 2000.

[210] R. P. Thomen, J. D. Quirk, D. Roach, T. Egan-Rojas, K. Ruppert, R. D. Yusen, T. A. Altes, D. A. Yablonskiy, and J. C. Woods. Direct comparison of 129Xe diffusion measurements with quantitative histology in human lungs. *Magnetic Resonance in Medicine*, 77(1):265–272, 2016.

[211] G. Tian, M. Hindle, S. Lee, and P. W. Longest. Validating CFD Predictions of Pharmaceutical Aerosol Deposition with in Vivo Data. *Pharmaceutical Research*, 32(10):3170–3187, 2015.

[212] A. W. Toga and P. M. Thompson. Mapping brain asymmetry. *Nature reviews. Neuroscience*, 4(1):37–48, 2003.

[213] N. J. Tustison, B. Contrella, T. a. Altes, B. B. Avants, E. E. de Lange, and J. P. Mugler. Longitudinal assessment of treatment effects on pulmonary ventilation using 1H/3He MRI multivariate templates. In J. B. Weaver and R. C. Molthen, editors, *SPIE Medical Imaging*, volume 8672, 2013.

[214] S. Ukil and J. M. Reinhardt. Anatomy-guided lung lobe segmentation and fissure analysis in x-ray CT images. *IEEE transactions on medical imaging*, 28(2):202–214, 2009.

[215] A. Valipour, P. L. Shah, W. Gesierich, R. Eberhardt, G. Snell, C. Strange, R. Barry, A. Gupta, E. Henne, S. Bandyopadhyay, P. Raffy, Y. Yin, J. Tschirren, and F. J. F. Herth. Patterns of Emphysema Heterogeneity. *Respiration*, 90(5):402–411, 2015.

[216] L. van der Maaten. Dimensionality Reduction Toolbox (https://lvdmaaten.github.io/drtoolbox/), 2017.

[217] B. van Ginneken. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiological Physics and Technology*, 10(1):1–10, 2017.

[218] B. Van Ginneken, W. Baggerman, and E. M. Van Rikxoort. Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5241 LNCS, pages 219–226. Springer, Berlin, Heidelberg, 2008.

[219] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE transactions on medical imaging*, 18(10):897–908, 1999.

[220] E. M. van Rikxoort, M. Prokop, B. de Hoop, M. A. Viergever, J. P. W. Pluim, and B. van Ginneken. Automatic segmentation of pulmonary lobes robust against incomplete fissures. *IEEE transactions on medical imaging*, 29(6):1286–96, 2010.

[221] E. M. Van Rikxoort, M. Prokop, B. De Hoop, M. a. Viergever, J. P. W. Pluim, and B. Van Ginneken. Automatic segmentation of pulmonary lobes robust against incomplete fissures. *IEEE Transactions on Medical Imaging*, 29(6):1286–1296, 2010.

[222] E. M. van Rikxoort and B. van Ginneken. Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review. *Physics in medicine and biology*, 58(17):R187–220, 2013.

[223] E. M. van Rikxoort, B. van Ginneken, M. Klik, and M. Prokop. Supervised enhancement filters: application to fissure detection in chest CT scans. *IEEE transactions on medical imaging*, 27(1):1–10, 2008.

[224] J. Vestbo, W. Anderson, H. O. Coxson, C. Crim, F. Dawber, L. Edwards, G. Hagan, K. Knobil, D. A. Lomas, W. MacNe, E. K. Silverman, and R. Tal-Singer. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *European Respiratory Journal*, 31(4):869–873, 2008.

[225] J. Vestbo, W. Anderson, H. O. Coxson, C. Crim, F. Dawber, L. Edwards, G. Hagan, K. Knobil, D. A. Lomas, W. MacNe, E. K. Silverman, and R. Tal-Singer. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *European Respiratory Journal*, 31(4):869–873, 2008.

[226] J. Vestbo, L. D. Edwards, P. D. Scanlon, J. C. Yates, A. Agusti, P. Bakke, P. M. A. Calverley, B. Celli, H. O. Coxson, C. Crim, D. A. Lomas, W. MacNee, B. E. Miller, E. K. Silverman, R. Tal-Singer, E. Wouters, S. I. Rennard, and ECLIPSE Investigators. Changes in forced expiratory volume in 1 second over time in COPD. *The New England journal of medicine*, 365(13):1184–92, 2011.

[227] J. Vestbo, S. S. Hurd, A. G. Agustí, P. W. Jones, A. A. Claus Vogelmeier, P. J. Barnes, Leonardo M. Fabbri, Fernando J. Martinez, Masaharu Nishimura, Robert A. Stockley, Don D. Sin, and R. Rodriguez-Roisin. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. *American journal of respiratory and critical care medicine*, 187(4):347–365, 2013.

[228] J. Vogel-Claussen, J. Renne, J. Hinrichs, C. Schönfeld, M. Gutberlet, F. Schaumann, C. Winkler, C. Faulenbach, N. Krug, F. K. Wacker, and J. M. Hohlfeld. Quantification of pulmonary inflammation after segmental allergen challenge using turbo-inversion recovery-magnitude magnetic resonance imaging. *American Journal of Respiratory and Critical Care Medicine*, 189(6):650–657, 2014.

[229] W.-d. Vogl, H. Prosch, M. Christina, U. Schmidt-erfurth, and G. Langs. Longitudinal Alignment of Disease Progression in Fibrosing Interstitial Lung Disease. In *Medical image computing and computer-assisted intervention (MICCAI)*, volume 7, pages 97–104, 2014.

[230] G. R. Washko. The role and potential of imaging in COPD. *Medical Clinics of North America*, 96(4):729–743, jul 2012.

[231] M. Weatherall, P. Shirtcliffe, J. Travers, and R. Beasley. Use of cluster analysis to define COPD phenotypes. *The European respiratory journal*, 36(3):472–4, 2010.

[232] M. Weatherall, J. Travers, P. M. Shirtcliffe, S. E. Marsh, M. V. Williams, M. R. Nowitz, S. Aldington, and R. Beasley. Distinct clinical phenotypes of airways disease defined by cluster analysis. *The European respiratory journal*, 34(4):812–8, 2009.

[233] W. Weder, R. Thurnheer, U. Stammberger, M. Bürge, E. W. Russi, and K. E. Bloch. Radiologic emphysema morphology is associated with outcome after surgical lung volume reduction. *Annals of Thoracic Surgery*, 64(2):313–320, 1997.

[234] J. A. Wedzicha and T. A. Seemungal. COPD exacerbations: defining their cause and prevention. *Lancet*, 370(9589):786–796, 2007.

[235] M. Weiger, M. Wu, M. C. Wurnig, D. Kenkel, W. Jungraithmayr, A. Boss, and K. P. Pruessmann. Rapid and robust pulmonary proton ZTE imaging in the mouse. *NMR in Biomedicine*, 27(9):1129–1134, 2014.

[236] J. M. Wells, G. R. Washko, M. K. Han, N. Abbas, H. Nath, a. J. Mamary, E. Regan, W. C. Bailey, F. J. Martinez, E. Westfall, T. H. Beaty, D. Curran-Everett, J. L. Curtis, J. E. Hokanson, D. a. Lynch, B. J. Make, J. D. Crapo, E. K. Silverman, R. P. Bowler, and M. T. Dransfield. Pulmonary arterial enlargement and acute exacerbations of COPD. *The New England journal of medicine*, 367(10):913–21, 2012.

[237] J. M. Wells, G. R. Washko, M. K. Han, N. Abbas, H. Nath, a. J. Mamary, E. Regan, W. C. Bailey, F. J. Martinez, E. Westfall, T. H. Beaty, D. Curran-Everett, J. L. Curtis, J. E. Hokanson, D. a. Lynch, B. J. Make, J. D. Crapo, E. K. Silverman, R. P. Bowler, and M. T. Dransfield. Pulmonary arterial enlargement and acute exacerbations of COPD. *The New England journal of medicine*, 367(10):913–21, 2012.

[238] R. Wiemker, T. Blow, and T. Blaffert. Unsupervised extraction of the pulmonary interlobar fissures from high resolution thoracic CT data. *International Congress Series*, 1281:1121–1126, 2005.

[239] J. M. Wild, H. Marshall, M. Bock, L. R. Schad, P. M. Jakob, M. Puderbach, F. Molinari, E. J. R. van Beek, and J. Biederer. MRI of the lung (1/3): Methods. *Insights into Imaging*, 3(4):345–353, 2012.

[240] V. Williams, M. Hardinge, S. Ryan, and A. Farmer. Patient's experience of identifying and managing exacerbations in COPD: A qualitative study. *Primary Care Respiratory Medicine*, 24(1):14062, nov 2014.

[241] J. C. Woods, C. K. Choong, D. A. Yablonskiy, J. Bentley, J. Wong, J. A. Pierce, J. D. Cooper, P. T. Macklem, M. S. Conradi, and J. C. Hogg. Hyperpolarized 3He diffusion MRI and histology in pulmonary emphysema. *Magnetic Resonance in Medicine*, 56(6):1293–1300, 2006.

[242] C. Xiao, B. Stoel, M. Bakker, Y. Peng, J. Stolk, and M. Staring. Pulmonary Fissure Detection in CT Images Using a Derivative of Stick Filter. *IEEE Transactions on Medical Imaging*, 35(6):1488–1500, 2016.

[243] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. a. Hoffman. MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. *IEEE transactions on medical imaging*, 25(4):464–75, 2006.

[244] D. A. Yablonskiy, A. L. Sukstanskii, J. C. Woods, D. S. Gierada, J. D. Quirk, J. C. Hogg, J. D. Cooper, and M. S. Conradi. Quantification of lung microstructure with hyperpolarized 3He diffusion MRI. *Journal of Applied Physiology*, 107(4):1258–1265, 2009.

[245] J. Yang, E. D. Angelini, B. M. Smith, J. H. M. Austin, E. A. Hoffman, D. A. Bluemke, R. G. Barr, and A. F. Laine. Explaining Radiological Emphysema Subtypes with Unsupervised Texture Prototypes: MESA COPD Study. In *MICCAI MCV*, volume 1, pages 1–10, 2016.

[246] C. Yilmaz, D. M. Dane, N. C. Patel, and C. C. W. Hsia. Quantifying Heterogeneity in Emphysema from High-Resolution Computed Tomography: A Lung Tissue Research Consortium Study. *Academic radiology*, pages 1–13, 2012.

[247] Y. Yin., E. A. Hoffman, and C.-L. Lin. Mass preserving nonrigid registration of CT lung images using cubic B-spline. *Medical physics*, 36(9):4213–22, 2009.

[248] J. Ying, J. Dutta, N. Guo, C. Hu, D. Zhou, A. Sitek, and Q. Li. Classification of Exacerbation Frequency in the COPDGene Cohort Using Deep Learning with Deep Belief Networks. *IEEE Journal of Biomedical and Health Informatics*, 2194(c):1–1, 2016.

[249] A. Young. *Modelling progression and heterogeneity in Alzheimer's disease*. PhD thesis, 2016.

[250] A. L. Young, R.-V. V. Marinescu, N. P. Oxtoby, M. Bocchetta, K. Yong, N. Firth, D. M. Cash, D. L. Thomas, K. M. Dick, J. Cardoso, J. van Swieten, B. Borroni, D. Galimberti, M. Masellis, M. C. Tartaglia, J. B. Rowe, C. Graff, F. Tagliavini, G. Frisoni, R. Laforce, E. Finger, A. Medonça, S. Sorbi, J. D. Warren, S. Crutch, N. C. Fox, S. Ourselin, J. M. Schott, J. D. Rohrer, D. C. Alexander, G. F. I. Genfi, and A. D. N. Initiative. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *bioRxiv*, page 236604, 2017.

[251] A. L. Young, N. P. Oxtoby, S. Ourselin, J. M. Schott, and D. C. Alexander. A simulation system for biomarker evolution in neurodegenerative disease. *Medical Image Analysis*, 26(1):47–56, 2015.

[252] P. A. Yushkevich, J. Piven, H. H. Cody, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31(3):1116–28, 2006.

[253] L. Zhang, E. A. Hoffman, and J. M. Reinhardt. Atlas-driven lung lobe segmentation in volumetric X-ray CT images. *IEEE transactions on medical imaging*, 25(1):1–16, 2006.

[254] L. Zhang, E. a. Hoffman, and J. M. Reinhardt. Atlas-driven lung lobe segmentation in volumetric X-ray CT images. *IEEE transactions on medical imaging*, 25(1):1–16, 2006.

[255] X. Zhou, T. Hayashi, T. Hara, H. Fujita, R. Yokoyama, T. Kiryu, and H. Hoshi. Automatic segmentation and recognition of anatomical lung structures from high-resolution chest CT images. *Computerized Medical Imaging and Graphics*, 30(5):299–313, 2006.